SMJERNICE OECD-A ZA DUBINSKU ANALIZU

ODGOVORNE UMJETNE INTELIGENCIJE

SMJERNICE OECD-A ZA DUBINSKU ANALIZU ODGOVORNE UMJETNE INTELIGENCIJE

HRVATSKA NACIONALNA
KONTAKTNA TOČKA

Nacionalna kontaktna točka za odgovorno poslovno ponašanje

RBCroatia.gov.hr
oecd-nkt@mingo.hr
nkt-oecd@mvep.hr

Ovo su djelo 26. siječnja 2026. odobrili Odbor za digitalnu politiku i Odbor za ulaganja te donijeli odluku o njegovoj javnoj objavi.

Izvorno objavio OECD na engleskom jeziku pod naslovom: OECD Due Diligence Guidance for Responsible AI @ OECD, 2026. (https://doi.org/10.1787/41671712-en).

Ovaj prijevod nije izradio OECD i ne bi ga se trebalo smatrati službenim prijevodom OECD-a. Za kvalitetu prijevoda i njegovu usklađenost s izvornim jezikom teksta odgovoran je isključivo autor ili autori prijevoda. U slučaju bilo kakvog odstupanja između izvornog djela i prijevoda, valjanim će se smatrati samo tekst izvornog djela.

Prijevod, lektura i korektura: Nina Matetić Pelikan (za ETNOtrend d.o.o. Samobor)

PREDGOVOR

Razvoj umjetne inteligencije ima potencijal preobraziti društvo na načine usporedive s industrijskom revolucijom ili pojavom interneta. Umjetna inteligencija ne predstavlja tek postupni napredak, već je transformativna tehnologija koja može povećati produktivnost, stvoriti gospodarsku vrijednost i riješiti složene izazove u različitim sektorima, kao što su zdravstvo, proizvodnja, logistika i javna uprava. Kako bi se iskoristio taj pozitivan potencijal, OECD utvrđuje uravnotežen pristup odgovornoj umjetnoj inteligenciji kojim se unapređuju mogućnosti koje ona pruža, istodobno pristupajući rizicima od štetnih učinaka.

Nakon donošenja revidirane Preporuke Vijeća o umjetnoj inteligenciji („Načela za umjetnu inteligenciju") u svibnju 2024., Vijeće OECD-a je na ministarskom sastanku zadužilo Odbor za digitalnu politiku („DPC") da u okviru Radne skupine za upravljanje umjetnom inteligencijom („AIGO") „nastavi svoj važan rad na području umjetne inteligencije nadovezujući se na ovu Preporuku" te „izradi i nastavi nadograđivati praktične smjernice za provedbu te Preporuke." Načela za umjetnu inteligenciju uviđaju pozitivan potencijal umjetne inteligencije da doprinese korisnim ishodima za ljude i planet. Načela za umjetnu inteligenciju promiču ekosustav za pouzdane sustave umjetne inteligencije utvrđivanjem načela i smjernica politika kojima se potiču inovacije uz istodobno rješavanje rizika.

Smjernice OECD-a za multinacionalna poduzeća o odgovornom poslovnom ponašanju („Smjernice za multinacionalna poduzeća"), ažurirane 2023., izričito navode da tehnološke inovacije potiču „produktivnost u svim sektorima, kao i sposobnost poduzeća da provode dubinsku analizu i doprinose održivom razvoju". Smjernice za multinacionalna poduzeća prepoznaju važnost širenja „utjecaja tehnološkog napretka na cijelo gospodarstvo, što uključuje rast produktivnosti i stvaranje radnih mjesta". Smjernice za multinacionalna poduzeća isto tako pozivaju poduzeća da provode dubinsku analizu odgovornog poslovnog ponašanja utemeljenu na riziku u pogledu stvarnih i potencijalnih štetnih učinaka povezanih sa znanošću, tehnologijom i inovacijama.

Namjena ovih Smjernica jest pomoći poduzećima u provedbi Smjernica za multinacionalna poduzeća i Načela za umjetnu inteligenciju. One služe kao instrument multinacionalnim poduzećima uključenima u lanac vrijednosti sustava umjetne inteligencije – onima koja

isporučuju ulazne elemente za razvoj umjetne inteligencije, aktivno sudjeluju u životnom ciklusu sustava umjetne inteligencije ili koriste sustave umjetne inteligencije u svojim poslovnim procesima, proizvodima i uslugama u svim sektorima.

Projekt zajednički nadziru Odbor za digitalnu politiku („DPC") putem Radne skupine za upravljanje umjetnom inteligencijom („AIGO") i Odbor za ulaganja („IC") putem Radne skupine za odgovorno poslovno ponašanje („WPRBC"). Ove se Smjernice temelje na radu Stručne skupine OECD.AI za rizik i odgovornost. Stručna skupina okuplja više od 100 predstavnika vlada, organizacija civilnog društva, predstavnika radnika te velikih i malih poduzeća iz cijelog lanca opskrbe umjetne inteligencije. Ove su Smjernice također u znatnoj mjeri poboljšane zahvaljujući pregledu i povratnim informacijama Savjetodavnog odbora civilnog društva za informacijsko društvo („CSISAC"), OECD Watcha, Savjetodavnog odbora sindikata („TUAC") i organizacije Business at OECD („BIAC").

SADRŽAJ

Okviri

SAŽETAK

Cilj je ovih Smjernica pomoći poduzećima u provedbi Smjernica za multinacionalna poduzeća i Načela za umjetnu inteligenciju. Ove su Smjernice zamišljene kao instrument koji mogu koristiti multinacionalna poduzeća uključena u lanac vrijednosti sustava umjetne inteligencije.

U prvom poglavlju uvodi se koncept dubinske analize odgovornog poslovnog ponašanja i donosi pregled šireg okvira politika upravljanja rizicima umjetne inteligencije. Također se opisuje ciljana publika i način na koji se ove Smjernice koriste kao instrument za snalaženje u okvirima upravljanja rizicima.

Smjernice se temelje na okviru OECD-a za dubinsku analizu, koji je opisan u Smjernicama za multinacionalna poduzeća i dodatno razrađen u Smjernicama OECD-a za dubinsku analizu odgovornog poslovnog ponašanja. Smjernice za multinacionalna poduzeća i povezani standardi OECD-a o odgovornom poslovnom ponašanju sadržavaju dobrovoljna načela odgovornog poslovnog ponašanja. Okvir za dubinsku analizu obuhvaća sljedeće mjere:

- 1. korak: ugraditi odgovorno poslovno ponašanje u politike i sustave upravljanja
- 2. korak: prepoznati i procijeniti stvarne i potencijalne štetne učinke
- 3. korak: zaustavljati, sprečavati i ublažavati štetne učinke
- 4. korak: pratiti provedbu i rezultate mjera dubinske analize
- 5. korak: komunicirati mjere poduzete radi rješavanja pitanja učinaka
- 6. korak: otklanjati učinke ili, u slučaju potrebe, surađivati na njihovu otklanjanju.

U 2. poglavlju izlaže se okvir dubinske analize odgovornog poslovnog ponašanja i primjeri njegove praktične provedbe za poduzeća uključena u razvoj i upotrebu sustava umjetne inteligencije. Okvir dubinske analize predstavljen u ovim Smjernicama na početku svakog koraka sadržava i pregled povezanih odredbi u postojećim okvirima, pri čemu se naznačuje na koji način svaki korak okvira dubinske analize dopunjavaju mjerodavne odredbe iz srodnih okvira za upravljanje rizicima umjetne inteligencije i kako je s njima povezan.

Uz svaki korak predlažu se „praktični primjeri provedbe" koji dodatno ilustriraju načine provedbe i, prema potrebi, prilagodbe popratnih

mjera i postupka dubinske analize. Praktični primjeri odabrani su tako da odgovaraju ovom kontekstu te se oslanjaju na vodeće okvire za upravljanje rizicima umjetne inteligencije, kao i na istraživanje dostupne literature i savjetovanja sa stručnjacima. Praktični primjeri nisu zamišljeni kao iscrpan popis. Neće svaki praktični primjer biti primjeren za svaku situaciju. Isto tako, u nekim situacijama poduzećima mogu biti korisni dodatni primjeri ili provedbene mjere.

Svrsishodnom provedbom preporuka postojećih okvira za upravljanje rizicima umjetne inteligencije, uključujući one opisane u ovim Smjernicama, poduzeća mogu udovoljiti mnogim očekivanjima pristupa dubinskoj analizi odgovornog poslovnog ponašanja. U nekim slučajevima okvir odgovornog poslovnog ponašanja pruža dodatna pojašnjenja i premošćuje praznine u drugim okvirima, osobito u pogledu uključivanja dionika i otklanjanja štetnih učinaka, koji su u postojećim okvirima obrađeni manje sveobuhvatno.

1

UVOD U DUBINSKU ANALIZU ODGOVORNOG POSLOVNOG PONAŠANJA I KLJUČNA RAZMATRANJA U PODRUČJU UMJETNE INTELIGENCIJE

—

U ovom se poglavlju uvodi koncept dubinske analize odgovornog poslovnog ponašanja i donosi pregled šireg okvira politika upravljanja rizicima umjetne inteligencije. Također se opisuje ciljana publika i način na koji se ove Smjernice koriste kao instrument za snalaženje u okvirima upravljanja rizicima.

UVOD U ODGOVORNU UMJETNU INTELIGENCIJU

Razvoj umjetne inteligencije ima potencijal preobraziti društvo na načine usporedive s industrijskom revolucijom ili pojavom interneta. Umjetna inteligencija ne predstavlja tek postupni napredak, već je transformativna tehnologija koja može povećati produktivnost, stvoriti gospodarsku vrijednost i riješiti složene izazove u različitim sektorima, kao što su zdravstvo, proizvodnja, logistika i javna uprava. Kako bi se iskoristio taj pozitivan potencijal, OECD utvrđuje uravnotežen pristup odgovornoj umjetnoj inteligenciji kojim se unapređuju mogućnosti koje umjetna inteligencija pruža i uspostavljaju uvjeti da ona bude profitabilnija, inovativnija i konkurentnija, istodobno pristupajući rizicima od štetnih učinaka.

Odgovorna umjetna inteligencija također ovisi o dobavljačima podataka, financiranju i fizičkoj infrastrukturi jednako kao i o digitalnim inovacijama. Smjernice OECD-a za multinacionalna poduzeća o odgovornom poslovnom ponašanju („Smjernice za multinacionalna poduzeća") (OECD, 2023[1]) i Preporuka OECD-a o umjetnoj inteligenciji („Preporuka OECD-a o UI-ju" ili, prema potrebi, „Načela za umjetnu inteligenciju") (OECD, 2024[2]) također naglašavaju ulogu svih poduzeća uključenih u razvoj i upotrebu sustava umjetne inteligencije. Takav pristup „na razini cijelog lanca vrijednosti" doprinijet će sigurnim i otpornim lancima vrijednosti umjetne inteligencije, otpornijima na poremećaje u opskrbnim lancima i na vanjske utjecaje.

Od ključne je važnosti da se razvoj i upotreba odgovorne umjetne inteligencije odvijaju uključujući dionike te stavljajući radnika u središte razmatranja, pri čemu se na tehnologiju gleda kao na nadopunu ljudskim sposobnostima, a ne kao na zamjenu za ljudski rad. Osiguravanjem smislenog uključivanja radnika i drugih dionika poduzeća mogu usmjeriti umjetnu inteligenciju prema primjenama koje nadopunjuju ljudski rad, umjesto da ga automatizacijom potiskuju.

Proaktivno pristupanje mogućim štetama povezanima sa sustavima umjetne inteligencije stvara temelj pouzdanosti koji može znatno ubrzati rast tržišta, kao i ulaganja. Predanost poduzećâ odgovornom razvoju i upotrebi umjetne inteligencije – kroz najbolje prakse opisane u ovim Smjernicama i drugim instrumentima OECD-a – omogućuje izgradnju povjerenja investitora, korisnika, regulatornih tijela i donositelja politika. To povjerenje postaje konkurentna prednost. Poduzeća koja steknu reputaciju odgovornog poslovanja lakše će pristupati tržištima kapitala, uspostavljati vrijedne poslovne odnose i snalaziti se u regulatornom okruženju. Odgovorna umjetna inteligencija ne samo

da ne koči inovacije, nego zapravo može ubrzati rast smanjenjem prepreka i sprečavanjem skupih reputacijskih i društvenih šteta koje bi inače mogle nastati.

S daljnjim razvojem međunarodnih regulatornih okvira, odgovorna i pouzdana umjetna inteligencija postaje sve važnija za pristup globalnim tržištima. Poduzeća koja proaktivno uključuju sprečavanje nastanka šteta u svoje postupke razvoja i uvođenja umjetne inteligencije u povoljnijem su položaju u pogledu prekograničnog širenja te pritom mogu izbjeći znatne troškove naknadne prilagodbe sustava radi usklađivanja s različitim regionalnim zahtjevima. Takav pristup usmjeren na budućnost može ono što bi se inače smatralo troškovima usklađivanja pretvoriti u strateška ulaganja koja donose povrat kroz širi pristup tržištima.

Odgovornu i pouzdanu umjetnu inteligenciju moguće je osobito uvjerljivo argumentirati s ekonomskog gledišta na temelju činjenice da poslovni korisnici u svoje postupke nabave sve češće uključuju upravljanje rizicima umjetne inteligencije, uslijed čega pouzdanost postaje ne samo etičko pitanje nego i poslovni zahtjev. Zahvaljujući tome, pristupanje pitanju šteta povezanih s umjetnom inteligencijom istodobno se vodi etičkim imperativima i poslovnim interesima te stvara pozitivan ciklus u kojem odgovorna inovacija potiče i društvenu korist i komercijalni uspjeh.

NAMJENA OVIH SMJERNICA

Cilj je ovih Smjernica pomoći poduzećima u provedbi Smjernica za multinacionalna poduzeća i Načela za umjetnu inteligenciju[1]. Ove su Smjernice zamišljene kao instrument kojim se mogu koristiti multinacionalna[2] poduzeća uključena u lanac vrijednosti sustava umjetne inteligencije (tj. poduzeća koja osiguravaju ulazne elemente za razvoj sustava umjetne inteligencije, imaju aktivnu ulogu u životnom ciklusu tih sustava ili ih koriste u svojim poslovnim procesima, proizvodima i uslugama u svim sektorima).

Ciljevi ovih Smjernica su:

- podržavati inovacije, ulaganja i rast poduzećâ u lancu vrijednosti umjetne inteligencije osiguravanjem jasnih uputa u pogledu načina na koji mogu proaktivno prepoznati stvarne i potencijalne štetne učinke (tj. rizike) koje mogu prouzročiti, kojima mogu doprinijeti ili s kojima mogu biti izravno povezana, odnosno pristupiti rješavanju problema tih učinaka, kao i iskoristiti pozitivne doprinose umjetne inteligencije društvu u područjima obuhvaćenima Smjernicama

za multinacionalna poduzeća i Načelima za umjetnu inteligenciju
- olakšati poduzećima snalaženje u postojećim međunarodnim i nacionalnim okvirima za upravljanje i rukovođenje rizicima umjetne inteligencije, kao i okvirima koji uključuju više dionika ili u kojima predvode pojedine gospodarske grane
- promicati usklađenost politika te, gdje je to moguće, interoperabilnost Smjernica za multinacionalna poduzeća, Načela za umjetnu inteligenciju i drugih nacionalnih ili međunarodnih okvira za upravljanje i rukovođenje rizicima umjetne inteligencije
- služiti kao zajednička referentna točka za okvire upravljanja rizicima umjetne inteligencije u različitim sustavima nadležnosti.

Smjernice se temelje na okviru OECD-a za dubinsku analizu koji je opisan u Smjernicama za multinacionalna poduzeća i dodatno razrađen u Smjernicama OECD-a za dubinsku analizu odgovornog poslovnog ponašanja („Smjernice za ODP") (OECD, 2018[3]). Smjernice za multinacionalna poduzeća i povezani standardi OECD-a o odgovornom poslovnom ponašanju sadržavaju dobrovoljna načela odgovornog poslovnog ponašanja. Preporuka vlada upućena poduzećima o tome da se pridržavaju Smjernica za multinacionalna poduzeća razlikuje se od pitanja pravne odgovornosti i primjene prava. Važno je napomenuti da su standardi OECD-a o odgovornom poslovnom ponašanju usklađeni s Vodećim načelima Ujedinjenih naroda o poslovanju i ljudskim pravima (Ured visokog povjerenika Ujedinjenih naroda za ljudska prava (OHCHR), 2012[4]) i Tripartitnom deklaracijom načela o multinacionalnim poduzećima i socijalnoj politici Međunarodne organizacije rada (ILO, 2023[5]) te ih nadopunjuju. Ovim se Smjernicama glavni okvir sadržan u Smjernicama za dubinsku analizu odgovornog poslovnog ponašanja nastoji pretočiti u konkretne i praktične mjere koje mogu poduzimati poduzeća kako bi prepoznala, spriječila, ublažila i otklonila stvarne i potencijalne štetne učinke povezane s razvojem i upotrebom sustava umjetne inteligencije.

Radi potpore globalnoj suradnji i usklađenosti politika za pouzdanu umjetnu inteligenciju te doprinosa interoperabilnosti gdje je to primjereno, ove se Smjernice oslanjaju na postojeće međunarodne i nacionalne okvire za upravljanje rizicima umjetne inteligencije, kao i propise i druge inicijative iz tog područja kako bi obuhvatile praktične primjere provedbe Smjernica za dubinsku analizu odgovornog poslovnog ponašanja u kontekstu umjetne inteligencije.

Osiguravajući poduzećima resurs za provedbu Smjernica za

multinacionalna poduzeća u praksi uz istodobnu dosljednost i usklađenost s Načelima za umjetnu inteligenciju, Smjernice nastoje osigurati podršku poduzećima koja posluju u više sustava nadležnosti i podliježu višestrukim regulatornim zahtjevima ili sudjeluju u više dobrovoljnih inicijativa kako bi ispunila ta očekivanja. To će pomoći poduzećima da zadrže povjerenje potrošača te im osigurati slobodu inovacije, kao i konkurentnost na globalnom tržištu.

CILJNA PUBLIKA

U Smjernicama za multinacionalna poduzeća poduzećima se preporučuje da provode dubinsku analizu u svrhu prepoznavanja i pristupanja problemu rješavanja bilo kakvih stvarnih i potencijalnih štetnih učinaka (tj. rizika) 1.) koje mogu prouzročiti ili kojima mogu doprinijeti svojim poslovanjem; ili 2.) kojima mogu doprinijeti ili s kojima mogu biti izravno povezana u okviru svojih poslovnih odnosa.

Ove se Smjernice u prvom redu obraćaju poduzećima koja pripadaju trima skupinama podrobnije opisanima u nastavku. Među njima su poduzeća uključena u životni ciklus sustava umjetne inteligencije[3] opisan u Načelima za umjetnu inteligenciju (planiranje i osmišljavanje, prikupljanje i obrada podataka, izgradnja modela i/ili njihova prilagodba, njihovo testiranje, evaluacija, verifikacija i validacija te uvođenje, kao i rad i praćenje sustava umjetne inteligencije). Smjernice su namijenjene i poduzećima uključenima u osiguravanje digitalnih, fizičkih i financijskih ulaznih elemenata za razvoj sustava umjetne inteligencije (npr. usluge anotacije podataka, pružatelji računalnih kapaciteta, pružatelji usluga računalstva u oblaku, proizvođači hardvera i investitori), kao i prodaju i licenciranje sustava umjetne inteligencije te trgovinu tim sustavima i njihova uvođenja, kako je opisano u Smjernicama za multinacionalna poduzeća i u skladu s revidiranom Preporukom o umjetnoj inteligenciji. To uključuje i poduzeća izvan „tehnološkog sektora" koja koriste sustave umjetne inteligencije u svojim poslovnim procesima, proizvodima i uslugama.

Iako bi okvir iz ovih Smjernica mogao biti mjerodavan i za razvoj i upotrebu drugog softvera i tehnologija, Smjernice se posebno usredotočuju na sustave umjetne inteligencije. Sustavi umjetne inteligencije jasno su definirani u Preporuci o umjetnoj inteligenciji i detaljno objašnjeni u pratećem memorandumu (OECD, 2024[6]).

Uloge u lancu vrijednosti umjetne inteligencije i poslovni odnosi između različitih aktera nisu linearni te se preklapaju. Isto tako, postupak razvoja umjetne inteligencije ne odvija se linearno. Primjerice, pružatelji

usluge razvojnog inženjeringa sustava umjetne inteligencije mogu ujedno biti i poduzeća koja projektiraju i proizvode hardver ili mogu sudjelovati u prikupljanju podataka i anotaciji skupova podataka. Radi razumijevanja dubinske analize odgovornog poslovnog ponašanja u kontekstu razvoja i upotrebe sustava umjetne inteligencije, Smjernice opisuju odgovornosti poduzeća u pogledu dubinske analize razvrstavanjem poduzeća u različite skupine prema aktivnosti koju obavljaju.[4] Stoga granice između triju skupina opisanih u nastavku nisu krute niti se skupine međusobno isključuju, već se njima nastoji poduzećima koja obavljaju različite aktivnosti pružiti informacije o pristupu dubinskoj analizi. Poduzeća mogu obavljati aktivnosti iz više skupina te bi trebala pristupati dubinskoj analizi u skladu s time. Isto tako, neka će poduzeća pri dubinskoj analizi prednost davati rizicima od štetnih učinaka koji proizlaze iz njihova poslovanja, dok će druga davati prednost rizicima koji proizlaze iz njihovih poslovnih odnosa.

Štetni učinci mogu biti povezani s aktivnostima opisanima u svakoj od skupina. Iako se većina primjera iz ovih Smjernica odnosi na pristupanje problemu štetnih učinaka povezanih s aktivnostima navedenima u skupini 2, od poduzeća iz skupina 1 i 3 se i dalje očekuje provedba dubinske analize kako bi pristupili problemu štetnih učinaka koje mogu prouzročiti svojim poslovanjem, proizvodima i uslugama.

Skupina 1: Dobavljači ulaznih elemenata za umjetnu inteligenciju

U ovu skupinu ubrajaju se poduzeća koja su dobavljači ulaznih elemenata za umjetnu inteligenciju *(AI inputs)*. Ona osiguravaju ulazne elemente za razvoj sustava umjetne inteligencije i općenito se smatraju početnim dijelom lanca vrijednosti sustava umjetne inteligencije. To uključuje aktivnosti povezane s osiguravanjem ulaznih elemenata u ekosustavu umjetne inteligencije (tj. vještina i resursa, kao što su podaci, kôd, algoritmi, modeli, istraživanja, znanje i iskustvo, programi osposobljavanja, rukovođenje, procesi te najbolje prakse potrebni za razumijevanje i sudjelovanje u životnom ciklusu sustava umjetne inteligencije, uključujući upravljanje rizicima). To uključuje poduzeća koja se bave sljedećim aktivnostima:

- pružanje i anotacija podataka
- izrada i kuriranje skupova podataka
- razvoj, prilagodba ili osiguravanje kôda za upotrebu trećih strana, uključujući doprinose otvorenim bibliotekama i softverskim komponentama za razvoj umjetne inteligencije
- razvoj pokazatelja i mjera evaluacije.

Također obuhvaća aktivnosti povezane s pružanjem financijskih, logističkih, administrativnih i hardverskih ulaznih elemenata potrebnih za potporu razvoju sustava umjetne inteligencije. To uključuje poduzeća koja se bave sljedećim aktivnostima:

- pružanje kapitala (npr. financijske institucije, društva rizičnog kapitala i drugi pružatelji kapitala)
- pružanje digitalne infrastrukture i administrativnih usluga (npr. pružatelji računalnih kapaciteta, pružatelji usluga računalstva u oblaku, platforme za digitalna plaćanja, digitalne platforme rada, operativni sustavi, trgovine aplikacijama, pružatelji sigurnosnog softvera i pružatelji poslovnog softvera)
- pružanje hardvera (npr. proizvođači i distributeri poluvodiča, dobavljači mrežne opreme i drugi proizvođači hardvera).

Ove Smjernice ne obuhvaćaju opskrbne lance hardverskih ulaznih elemenata (npr. rudarenje sirovina i proizvodnju hardverskih komponenti), koji su predmet zasebnih smjernica za dubinsku analizu odgovornog poslovnog ponašanja.[5]

Skupina 2: Poduzeća aktivno uključena u projektiranje, razvoj, uvođenje i rad sustava umjetne inteligencije

U ovu skupinu ubrajaju se poduzeća koja su uključena u aktivnosti životnog ciklusa sustava umjetne inteligencije navedene u nastavku. Razumijevanje životnog ciklusa sustava umjetne inteligencije može pomoći svim poduzećima da bolje prepoznaju rizike, kao i da na odgovarajući način pristupe interakciji u okviru poslovnih odnosa s poduzećima iz skupine 2. Među njima su poduzeća koja se bave sljedećim aktivnostima:

- planiranje i projektiranje sustava
- izgradnja modela i/ili prilagodba postojećeg modela
- testiranje, evaluacija, verifikacija i validacija modela i sustava
- uvođenje[6] sustava neovisno o kanalu distribucije (uključujući distribuciju softvera otvorenog kôda)
- upravljanje sustavom za korisnike i praćenje njegova rada.

Skupina 2 može obuhvaćati i poduzeća koja prilagođavaju i ponovno uvode postojeće modele umjetne inteligencije za slučajeve korištenja specifične za pojedino poduzeće.

Skupina 3: Korisnici sustava umjetne inteligencije

Poduzeća iz ove skupine koriste sustave umjetne inteligencije u svojim poslovnim procesima, proizvodima i uslugama te se općenito smatraju završnim dijelom lanca vrijednosti sustava umjetne inteligencije. Među njima su financijske institucije i poduzeća u „realnom gospodarstvu" (tj. proizvođači i prodavatelji robe i usluga, uključujući one koji nisu povezani sa sustavima ili tehnologijom umjetne inteligencije). Poduzeća iz ove skupine trebala bi kao dio šireg postupka dubinske analize u svim svojim poslovnim procesima i odnosima razmotriti provedbu dubinske analize sustava umjetne inteligencije čiji su korisnici. To znači da bi trebala davati prednost rizicima od štetnih učinaka koje predstavlja sustav umjetne inteligencije u odnosu na druge rizike koje poduzeće može prouzročiti, kojima može doprinijeti ili s kojima može biti povezano u svojim sektorima. Primjerice, ako sustav umjetne inteligencije nije povezan sa znatnim rizicima, poduzeće može dati prednost drugim temama odgovornog poslovnog ponašanja za djelovanje, uključujući teme koje nisu povezane sa sustavima umjetne inteligencije.

OKVIR 1.1. RAZMATRANJA ZA MALA I SREDNJA PODUZEĆA (MSP-OVE)

Sustavi umjetne inteligencije imaju potencijal donijeti MSP-ovima znatne gospodarske koristi, među ostalim kroz pristup alatima koji mogu omogućiti veću učinkovitost uz niže troškove. Osim toga, imaju ključnu ulogu u više faza razvoja sustava umjetne inteligencije. U skladu sa standardima odgovornog poslovnog ponašanja, od MSP-ova se, kao i od drugih poduzeća, očekuje provedba dubinske analize. Standardi odgovornog poslovnog ponašanja ipak priznaju da MSP-ovi možda nemaju iste kapacitete za ispunjavanje očekivanja povezanih s dubinskom analizom kao veća poduzeća. MSP-ovi koji su u početnoj fazi istraživanja i razvoja, dokazivanja koncepta i financiranja djeluju s ograničenim resursima te ih obično usmjeravaju na neposrednije, praktične potrebe komercijalizacije svojih proizvoda ili usluga. Stoga MSP-ovi općenito mogu biti suočeni s izazovima u provedbi odgovornog poslovnog ponašanja povezanima s uključivanjem dionika,

ostvarivanjem utjecaja na poslovne odnose i pokrivanjem troškova potrebnih za poduzimanje mjera sprečavanja i ublažavanja rizika. Radi prevladavanja tih izazova i promicanja provedbe standarda odgovornog poslovnog ponašanja, ove Smjernice osobito potiču MSP-ove da, kad god je to moguće, primjenjuju suradničke pristupe i sudjeluju u strukovnim inicijativama radi udruživanja resursa (u skladu s pravilima tržišnog natjecanja), smanjenja troškova dubinske analize te olakšavanja pristupa informacijama o rizicima od štetnih učinaka i njihova usklađivanja. Standardi odgovornog poslovnog ponašanja također priznaju da priroda i opseg dubinske analize trebaju biti razmjerni veličini poduzeća, njegovoj uključenosti u štetni učinak i ozbiljnosti štetnog učinka.

Uvažavajući te okolnosti, Smjernicama za multinacionalna poduzeća nastoji se osigurati da MSP-ovi ne budu izloženi nerazmjernim opterećenjima te im omogućiti da se usmjere na najrelevantnije rizike u okviru svojih kapaciteta. Nadalje, potiču veća poduzeća da prednost daju stranama u poslovnim odnosima koji su MSP-ovi kako bi podržali njihove postupke dubinske analize.

Osim što će MSP-ovi, ispunjavajući standarde odgovornog poslovnog ponašanja, ispuniti međunarodna očekivanja, to im može otvoriti nova tržišta ili omogućiti bolji pristup financiranju. Može pomoći pri zapošljavanju ili zadržavanju zaposlenika. Isto tako, može se pokazati ključnim za uključivanje u lance vrijednosti s obzirom na to da se veći poslovni partneri sve češće suočavaju s praksama dubinske analize odgovornog poslovnog ponašanja.

MSP-ovi mogu iskoristiti mreže suradnje (npr. regionalne inicijative za umjetnu inteligenciju i digitalnu transformaciju[1] te regionalne inicijative odgovornog poslovnog ponašanja[2]) kako bi pri provedbi standarda odgovornog poslovnog ponašanja i umjetne inteligencije dobili dodatnu tehničku podršku i pojašnjenja.

Bilješke:
1. V., npr., dijalog OECD-a i Afričke unije o umjetnoj inteligenciji.
2. V., npr., programe globalnog angažmana OECD-a u Aziji, regiji Bliskog istoka i sjeverne Afrike (MENA) i Latinskoj Americi.

Ostale mjerodavne ciljne skupine

Smjernice za multinacionalna poduzeća imaju i jedinstveni mehanizam za promicanje Smjernica i podnošenje pritužbi, koji se oslanja na nacionalne kontaktne točke za odgovorno poslovno ponašanje

(NKT-ove).[7] Ove Smjernice mogu biti koristan resurs i NKT-ovima u promicanju Smjernica za multinacionalna poduzeća i donošenju odluka povezanih s odgovornošću za navodne povrede tih Smjernica.

Ove Smjernice mogu biti korisne i pri izradi standarda povezanih s odgovornom umjetnom inteligencijom, primjerice donositeljima politika, regulatornim tijelima te strukovnim inicijativama i inicijativama koje uključuju više dionika u svrhu usklađivanja s međunarodnim standardima.

Ostale mjerodavne ciljne skupine mogu uključivati organizacije civilnog društva, radnike, predstavnike radnika, sindikate, strukovna udruženja i nacionalna regulatorna tijela, uključujući tijela za zaštitu podataka i sektorska nadzorna tijela.

Osobitu pozornost treba posvetiti izazovima u provedbi s kojima se suočavaju poduzeća i vlade u zemljama u razvoju, uključujući potrebu za jačanjem kapaciteta, tehničkom pomoći i diferenciranim smjernicama prilagođenima lokalnim regulatornim i institucionalnim okolnostima.

Naposljetku, ove su Smjernice relevantne i za pojedince i skupine te njihove predstavnike koji su bili ili bi mogli biti nepovoljno pogođeni sustavom umjetne inteligencije.

RAZUMIJEVANJE RIZIKA POVEZANIH S RAZVOJEM I UPOTREBOM UMJETNE INTELIGENCIJE

Razvoj i upotreba sustava umjetne inteligencije mogli bi pozitivno utjecati na pitanja obuhvaćena Smjernicama za multinacionalna poduzeća. Primjerice, upotreba sustava umjetne inteligencije može omogućiti znatna poboljšanja u području zaštite zdravlja i sigurnosti na radu automatizacijom opasnih zadataka. U javnoj upravi korištenje umjetne inteligencije u pametnim mrežama, pametnim gradovima i povezanim uređajima može pomoći u predviđanju potreba za održavanjem infrastrukture i usmjeravanju prometnih tokova radi smanjenja prometnih zagušenja. Sposobnost umjetne inteligencije da brzo analizira goleme količine podataka, prepoznaje obrasce i izgrađuje prediktivne modele čini je važnim alatom za otkrivanje financijskog kriminala, borbu protiv otmica i trgovine ljudima, prepoznavanje situacija dužničkog ili dječjeg rada te analizu mjesta počinjenja kaznenih djela. Šire gledano, upotreba sustava umjetne inteligencije pruža mogućnosti za inovacije, gospodarski rast i promicanje ljudskih prava. Kako bi se ostvarile te pozitivne koristi, važno je učinkovito upravljati rizicima od štetnih učinaka povezanima sa sustavima umjetne

inteligencije.[8] Pri razmatranju cijelog lanca vrijednosti sustava umjetne inteligencije može biti mjerodavan širi raspon rizika. Tako je dokazan učinak znatne računalne snage koja se koristi za osposobljavanje i upotrebu nekih vrsta sustava umjetne inteligencije (OECD, 2022[7]), dok su neke usluge koje obavljaju ljudi, kao što su usluge obogaćivanja podataka, dovele do štetnih radnih praksi (Partnership on AI, 2021[8]). S druge strane, baš kao što je slučaj s mnogim novim tehnologijama, zlonamjerni akteri iz javnog i privatnog sektora mogu pronaći načine za zloupotrebu sustava umjetne inteligencije. Značajan dvojaki potencijal sustavâ umjetne inteligencije i mogućnosti njihove prenamjene mogu voditi štetnoj upotrebi čak i ako je namjena za koju su projektirani bezazlena.

Smjernice za multinacionalna poduzeća prepoznaju potrebu za time da poduzeća provode dubinsku analizu utemeljenu na riziku u pogledu stvarnih i potencijalnih štetnih učinaka svojih aktivnosti povezanih s tehnološkim inovacijama. Isto tako prepoznaju i potrebu za time da poduzeća uključena u razvoj novih tehnologija ili novih primjena postojećih alata predvide štetne učinke i izazove koje tehnologije nameću, istodobno promičući odgovorne inovacije.

Na međunarodnoj, regionalnoj i nacionalnoj razini postoje brojni okviri koji opisuju rizike povezane s razvojem i upotrebom sustava umjetne inteligencije te preporučuju mjere koje bi poduzeća trebala poduzeti radi pristupanja problemu tih rizika. Opseg rizika obuhvaćenih tim okvirima varira. Iako popis okvira može poslužiti kao polazna točka za prepoznavanje niza potencijalno mjerodavnih rizika u okviru dubinske analize pojedinog poduzeća, on nije iscrpan, a mnogi se rizici preklapaju i mogu biti međusobno povezani. Isto tako, nisu svi okviri mjerodavni za svako poduzeće. Od svakog se poduzeća očekuje da utvrdi svoja prioritetna područja rizika na temelju okolnosti u kojima posluje, uključujući i dodatne rizike koji nisu na popisu. Ove Smjernice zauzimaju pristup koji nije usmjeren na pojedine vrste rizika kako bi ostale trajno primjenjive. S daljnjim razvojem političkih stajališta i dubljim razumijevanjem rizika povezanih sa sustavima umjetne inteligencije, kao dopuna ovim Smjernicama koristit će se i rezultati budućih istraživanja.

OBILJEŽJA POUZDANE UMJETNE INTELIGENCIJE

Namjera je ovih Smjernica isto tako i omogućavati odgovorno upravljanje te razvoj pouzdanih sustava umjetne inteligencije. U kontekstu Smjernica pojam „pouzdana umjetna inteligencija” odnosi se na

sustave umjetne inteligencije koji utjelovljuju Načela OECD-a za umjetnu inteligenciju, ažurirana 2024. (OECD, 2024[9]). Ta načela proizašla su iz nalaza na temelju kojih se procjenjuju rizici i utvrđuju odgovornosti za ublažavanje i sprečavanje štetnih učinaka.

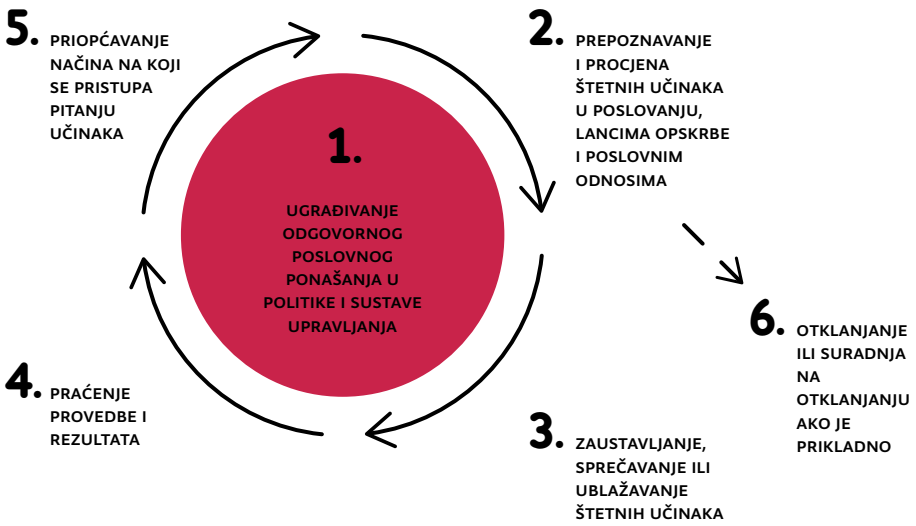
OSNOVE DUBINSKE ANALIZE ODGOVORNOG POSLOVNOG PONAŠANJA

Okvir za dubinsku analizu odgovornog poslovnog ponašanja

Smjernice za multinacionalna poduzeća utvrđuju dobrovoljni okvir dubinske analize za poduzeća koji se vlade obvezuju aktivno promicati i provoditi. Njime su utvrđene sljedeće mjere:

1. ugrađivanje odgovornog poslovnog ponašanja u politike i sustave upravljanja
2. prepoznavanje i procjenjivanje stvarnih i potencijalnih štetnih učinaka povezanih s poslovanjem, proizvodima i uslugama poduzeća
3. zaustavljanje, sprečavanje i ublažavanje štetnih učinaka
4. praćenje provedbe i rezultata
5. priopćavanje načina na koji se pristupa pitanju učinaka
6. osiguravanje otklanjanja ili surađivanje na otklanjanju kad god je to moguće.

SLIKA 1.1. GRAFIČKI PRIKAZ OKVIRA ZA DUBINSKU ANALIZU ODGOVORNOG POSLOVNOG PONAŠANJA



Izvor: OECD (2018[3]), *Smjernice OECD-a za dubinsku analizu odgovornog poslovnog ponašanja*, https://rbcroatia.gov.hr/wp-content/uploads/2024/04/Smjernice-OECD-a-za-dubinsku-analizu-1.pdf.

Zamišljeno je da se ti koraci poduzimaju istodobno te da ih se ponavlja jer je dubinska analiza kontinuirani proces, istodobno proaktivan i reaktivan. Koraci su opisani detaljnije te u kontekstu razvoja i upotrebe umjetne inteligencije u 2. poglavlju ovih Smjernica.

Okvir za dubinsku analizu odgovornog poslovnog ponašanja u velikoj je mjeri usklađen s drugim okvirima za upravljanje rizicima umjetne inteligencije (OECD, 2023[10]) od kojih se mnogi u velikoj mjeri međusobno preklapaju u pogledu ključnih pitanja. Svaki korak iz ovih Smjernica izravno upućuje na povezane zahtjeve u postojećim okvirima za upravljanje rizicima umjetne inteligencije te stoga može poslužiti kao pomoć pri upućivanju na različite izvore te pri usklađenoj provedbi zahtjeva u različitim okvirima i sustavima nadležnosti. Svrsishodnom provedbom preporuka drugih postojećih okvira za upravljanje rizicima umjetne inteligencije poduzeća mogu udovoljiti mnogim očekivanjima koja proizlaze iz pristupa dubinskoj analizi odgovornog poslovnog ponašanja. U nekim slučajevima okvir odgovornog poslovnog ponašanja pruža dodatnu jasnoću i premošćuje praznine u drugim okvirima, osobito u pogledu uključivanja dionika i otklanjanja štetnih učinaka, koji su u postojećim okvirima obrađeni manje sveobuhvatno.

Odnos koji podrazumijeva pravne obveze

Smjernice za multinacionalna poduzeća sadržavaju dobrovoljna načela i standarde odgovornog poslovnog ponašanja koji su u skladu s primjenjivim pravom i međunarodno priznatim standardima. Pitanja obuhvaćena Smjernicama za multinacionalna poduzeća mogu biti predmet nacionalnog zakonodavstva i međunarodnih obveza. Smjernice za multinacionalna poduzeća sadržavaju preporuke o odgovornom poslovnom ponašanju koje mogu nadilaziti pravne obveze poduzeća. Preporuka vlada upućena poduzećima da se pridržavaju Smjernica za multinacionalna poduzeća razlikuje se od pitanja pravne odgovornosti i primjene prava (v. Smjernice za multinacionalna poduzeća, Predgovor, t. 5. (OECD, 2023[1])).

U Smjernicama za multinacionalna poduzeća navodi se kako su poduzeća u prvom redu obvezna poštovati nacionalno pravo sustava nadležnosti u kojima posluju i/ili imaju poslovni nastan (v. Smjernice za multinacionalna poduzeća, pogl. 1., t. 2. (OECD, 2023[1])). Dubinska analiza može pomoći poduzećima u ispunjavaju njihovih pravnih obveza u pogledu pitanja koja se odnose na odgovorno poslovno ponašanje. U sustavima nadležnosti u kojima su nacionalni zakoni i propisi u

sukobu s načelima i standardima Smjernica za multinacionalna poduzeća dubinska analiza također može pomoći poduzećima da se u najvećoj mogućoj mjeri pridržavaju Smjernica za multinacionalna poduzeća. Isto tako, nacionalno pravo može u nekim slučajevima zahtijevati od poduzeća da poduzme mjere u vezi s određenim pitanjem odgovornog poslovnog ponašanja (primjerice zakonima koji se odnose na konkretna tematska područja odgovornog poslovnog ponašanja, kao što su internetski rizici za maloljetnike).

Očekivanja u pogledu dubinske analize koja proizlaze iz Smjernica za multinacionalna poduzeća ili na njih upućuju sve se više ugrađuju u pravne zahtjeve. Iako ove Smjernice mogu pomoći poduzećima i vladama da bolje razumiju kako bi mogle provoditi neke od tih pravnih zahtjeva, na njih se ne treba oslanjati kao na jedinstveni model za postizanje usklađenosti.

Povjerljivost poslovnih podataka

Pri provedbi dubinske analize odgovornog poslovnog ponašanja potrebno je posvetiti dovoljnu pozornost povjerljivosti poslovnih podataka, poslovnim tajnama, poslovno osjetljivim informacijama i mogućim ograničenjima iz prava tržišnog natjecanja povezanima s razmjenom takvih informacija, kao i informacijama zaštićenima pravima intelektualnog vlasništva. Iako su to legitimne prepreke nekim aspektima objavljivanja informacija, transparentnosti i uključivanja dionika, od poduzeća se ipak očekuje da u dobroj vjeri nastoje priopćavati informacije i smisleno uključivati dionike, pritom na odgovarajući način uvažavajući povjerljivost, pravo tržišnog natjecanja, druga relevantna pravna pitanja te legitimne razloge za povjerljivost.

Provedba odgovornog poslovnog ponašanja u okvirima prava tržišnog natjecanja

Suradnja s konkurentima ili stranama u poslovnim odnosima radi podupiranja provedbe odgovornog poslovnog ponašanja, uključujući u okviru inicijativa održivosti, podliježe pravu tržišnog natjecanja (OECD, 2015[11]).

Prema Smjernicama za multinacionalna poduzeća, iako bi poduzeća i inicijative za suradnju u kojima sudjeluju trebali poduzeti proaktivne mjere kako bi razumjeli pitanja prava tržišnog natjecanja u svojoj nadležnosti te izbjegli aktivnosti koje bi mogle predstavljati povredu prava tržišnog natjecanja, vjerodostojne inicijative za odgovorno poslovno ponašanje nisu same po sebi u sukobu s ciljevima prava

tržišnog natjecanja te suradnja u okviru takvih inicijativa u pravilu ne predstavlja povredu takvih propisa (v. Smjernice za multinacionalna poduzeća, pogl. X., t. 121. (OECD, 2023[1])).

Postoje tri opće praktične mjere koje poduzeća mogu razmotriti radi razumijevanja pitanja povezanih sa suradnjom i pravom tržišnog natjecanja:

- traženje savjeta od tijela nadležnih za tržišno natjecanje: ako nisu sigurna bi li se određeno ponašanje ili oblik suradnje mogli smatrati protivnima pravu tržišnog natjecanja te time predstavljati regulatorni rizik, poduzeća mogu zatražiti savjet tijela nadležnih za tržišno natjecanje
- prakticiranje transparentnosti: nadležna tijela obično su skeptičnija prema inicijativama ili sporazumima između konkurenata ako je ponašanje potpuno netransparentno. Stoga transparentnost u vezi s inicijativama odgovornog poslovnog ponašanja može biti koristan način za ublažavanje zabrinutosti u pogledu tržišnog natjecanja. Važno je naglasiti da sama činjenica da je sporazum javan ili da postoji transparentnost oko neke inicijative ne isključuje primjenu prava ako je riječ o protutržišnom ponašanju. Međutim, transparentnost može doprinijeti rasvjetljavanju potencijalno problematičnih pitanja i time osigurati njihovo brzo rješavanje
- ugrađivanje inicijativa odgovornog poslovnog ponašanja u programe usklađenosti: budući da su poduzeća odgovorna za samoprocjenu svojeg ponašanja u pogledu prava tržišnog natjecanja, potiče ih se da izrađuju i provode programe usklađenosti koji će ukazivati na postojanje rizika i doprinijeti razumijevanju načina upravljanja rizicima na razini organizacije. Većina velikih poduzeća vjerojatno već ima uspostavljene programe usklađenosti s pravilima tržišnog natjecanja koji se mogu koristiti kao referenca ili prilagoditi za potrebe konkretnih inicijativa suradnje u području odgovornog poslovnog ponašanja.

Svrhovito uključivanje dionika

Svrhovito[9] uključivanje dionika, osobito radnika, predstavnika radnika i sindikata, zajednica izloženih mogućim štetnim učincima ili drugih dionika koji su najizloženiji rizicima štetnih učinaka, ključno je za učinkovitu dubinsku analizu. Takvo uključivanje doprinosi razvoju pouzdanih sustava umjetne inteligencije. Uključivanje dionika sastavni je dio svih koraka okvira za dubinsku analizu. U nekim sustavima

nadležnosti uključivanje dionika može biti i zasebno pravo (npr. pristanak pacijenta pri primjeni umjetne inteligencije u medicinskim kontekstima, v. i Akt EU-a o UI-ju, čl. 61.).

Svrhovito uključivanje dionika može imati i brojne koristi za poduzeća, među ostalim izgradnju povjerenja i otpornosti na krize te snažnije usklađivanje s tržišnim i društvenim očekivanjima. Ako su dionici uključeni tijekom cjelokupnog postupka dubinske analize, razumiju ne samo koje su odluke donesene nego i zašto, čime se stvara povjerenje u postupak čak i kad se ne slažu sa svakom odlukom. Vanjsko savjetovanje u ranoj fazi omogućava prilagodbu pristupa prije ulaganja značajnih resursa u rješavanje manje značajnih rizika, čime se mogu smanjiti troškovi.

Perspektive vanjskih dionika mogu pomoći u prepoznavanju tržišnih segmenata koji nisu dovoljno obuhvaćeni ili slučajeva korištenja koji su zanemareni. Isto tako, izravno uključivanje krajnjih korisnika sustava umjetne inteligencije može otkriti stvarne, a ne pretpostavljene potrebe. Primjerice, poduzeća koja razvijaju i koriste sustave umjetne inteligencije u zdravstvu mogu se savjetovati s medicinskim stručnjacima i predstavnicima određenih skupina pacijenata te uvidjeti da je mogućnost interpretacije važnija od marginalnih poboljšanja točnosti. Sustavi umjetne inteligencije razvijeni uz doprinos dionika se, osim toga, suočavaju s manje prepreka pri prihvaćanju jer se proaktivno pristupilo rješavanju ključnih pitanja.

Dionici mogu biti izloženi mogućim štetnim učincima u više faza razvoja i upotrebe sustava umjetne inteligencije. To, primjerice, uključuje pojedince čiji se osobni podaci ili intelektualno vlasništvo upotrebljavaju za treniranje sustava umjetne inteligencije, radnike uključene u usluge obogaćivanja podataka u fazi razvoja te zajednice izložene mogućim štetnim učincima povezanima s računalnim kapacitetima umjetne inteligencije. Nakon uvođenja može, primjerice, uključivati radnike koje nadziru sustavi umjetne inteligencije i pojedince na koje utječu javne usluge u kojima se koriste sustavi umjetne inteligencije.

Uključivanje dionika obuhvaća interaktivne postupke uključivanja mjerodavnih dionika, primjerice putem sastanaka, saslušanja ili savjetodavnih postupaka. Relevantni dionici su osobe ili skupine ili njihovi ovlašteni predstavnici koji imaju prava ili interese u vezi s područjima obuhvaćenima Smjernicama na koje utječu ili bi mogli utjecati štetni učinci povezani s aktivnostima poduzeća počevši od razvoja, preko uvođenja, rada, financiranja, prodaje, licenciranja, do trgovine i/ili upotrebe sustava umjetne inteligencije.

Kako bi uključivanje dionika bilo svrhovito, treba biti dvosmjerno, provoditi se u dobroj vjeri i uzimati u obzir stavove dionika. Dionicima treba pravodobno pružiti istinite i potpune informacije te im omogućiti da iznesu svoja stajališta prije donošenja važnih odluka koje bi mogle utjecati na njih. Prema potrebi, osobito je važno da dionici aktivno sudjeluju u prepoznavanju štetnih učinaka.

Uslijed brzog razvoja i izmjena sustava umjetne inteligencije u stvarnom vremenu, poduzeća će možda trebati osmisliti ili prilagoditi postojeće prakse kako bi osigurala svrhovito uključivanje dionika. Na uključivanje dionika ne bi se trebalo gledati kao na jednokratan događaj, nego kao na kontinuiran proces ugrađen u životni ciklus sustava umjetne inteligencije te, prema potrebi, i u druge aspekte razvoja i upotrebe umjetne inteligencije (npr. pri prikupljanju podataka ili tijekom krajnje upotrebe). U praksi postoji niz načina na koje poduzeća mogu uključivati dionike.[10] Dionici mogu sudjelovati:

- u internim raspravama o namjenama proizvoda i željenom učinku
- u oblikovanju proizvoda
- u pripremi i validaciji skupova podataka
- u treniranju i testiranju
- kontinuirano tijekom upotrebe sustava umjetne inteligencije
- u testiranju i evaluaciji sustava umjetne inteligencije nakon uvođenja
- kroz inicijative koje uključuju više dionika i postupke neovisne procjene
- u redovitim osposobljavanjima radnika u kontekstima u kojima se sustavi umjetne inteligencije koriste za upravljanje aktivnostima radnika. Ključno je da se radnike redovito obavještava o mogućnostima i rizicima sustava umjetne inteligencije kako bi mogli svrhovito sudjelovati u raspravama o postupku dubinske analize.

Uz sve druge mehanizme koje mogu uspostaviti u vezi s uključivanjem dionika, poduzeća trebaju poštovati pravo radnika da osnuju sindikate i predstavničke organizacije po vlastitom izboru ili im se pridruže, među ostalim tako da se suzdrže od uplitanja u izbor radnika da, po svom izboru, osnuju sindikat ili predstavničku organizaciju ili im se pridruže (OECD, 2023., Smjernice za multinacionalna poduzeća, pogl. V., t. 1.(a)).

U vrlo velikim poduzećima, koja mogu razvijati ili koristiti stotine sustava umjetne inteligencije, možda neće biti izvedivo uključivanje dionika u više faza razvoja svakog pojedinog sustava. Mala i srednja

poduzeća mogu se suočiti i s izazovima povezanima s resursima i pristupom pri uključivanju dionika. Sva se poduzeća mogu suočiti s izazovima povezanima s brzinom razvoja umjetne inteligencije, što može ograničiti dostupnost mjerodavnih dionika koje bi trebalo uključiti. Ti izazovi pokazuju kako uključivanje dionika zahtijeva pažljivo planiranje i kako poduzeća mogu birati između različitih oblika uključivanja dionika, među kojima je i pristup na višoj razini, s ciljem prenošenja stečenih spoznaja na razinu proizvoda. Uključivanje na razini pojedinog proizvoda (npr. uključivanje u fazi oblikovanja ili uključivanje dionika izloženih štetnim učincima) možda će biti izvedivo samo u određenim visokorizičnim kontekstima.

Izazov povezan s uključivanjem može predstavljati i ograničena razina znanja dionika o umjetnoj inteligenciji i tematskim područjima odgovornog poslovnog ponašanja. Kad dionici razumiju mogućnosti, ograničenja i potencijalne posljedice sustava umjetne inteligencije, mogu donositi informirane odluke i pružati svrhovit doprinos postupcima upravljanja rizicima. To im znanje omogućava da prepoznaju potencijalne štetne učinke prije nego što nastanu i zauzmu se za odgovorne prakse razvoja i uvođenja sustava. Ulaganjem u edukaciju i transparentnu komunikaciju o sustavima umjetne inteligencije poduzeća mogu izgraditi povjerenje i potaknuti zajedničko rješavanje problema, što u konačnici dovodi do učinkovitijeg postupka dubinske analize.

Poduzeća i dionike potiče se da zajedno prepoznaju metode uključivanja koje su za njih izvedive i učinkovite. Poduzeća trebaju dati prednost uključivanju onih dionika, ili njihovih posrednika, za koje postoji najveća vjerojatnost da će na njih utjecati aktivnosti poduzeća. Osobito treba nastojati uključivati dionike koji su najizloženiji rizicima štetnih učinaka.

KAKO KORISTITI OVE SMJERNICE

Cilj je ovih Smjernica pružiti okvir za potporu poduzećima pri provedbi Smjernica za multinacionalna poduzeća, Smjernica za dubinsku analizu odgovornog poslovnog ponašanja i Preporuke OECD-a o umjetnoj inteligenciji te ih treba koristiti zajedno s tim standardima, kao i s drugim međunarodnim, nacionalnim i sektorskim okvirima, inicijativama i drugim izvorima informacija o upravljanju rizicima, osobito smjernicama namijenjenima konkretnim kontekstima koje pružaju detaljnije informacije o određenim rizicima ili slučajevima korištenja. U dokumentu se upućuje i na druge izvore informacija.

Uzimajući u obzir da se mjerodavni propisi i dobrovoljni okviri

povezani s umjetnom inteligencijom razlikuju od države do države, poduzeća mogu prilagoditi svoje mjere dubinske analize konkretnim okolnostima i regulatornim okruženjima u kojima posluju. Osim toga, usklađenost s tim propisima ili okvirima često će doprinijeti poštovanju povezanih odredbi ovih Smjernica.

Prije no što se okrenu izvorima namijenjenima konkretnim kontekstima na koje se upućuje u bilješkama ili modulima dostupnima u OECD.AI katalogu alata i pokazatelja, korisnici ovih Smjernica mogu pročitati i razumjeti osnovni okvir i praktične primjere (OECD, bez datuma[12]).

2

OKVIR ZA DUBINSKU ANALIZU I PRAKTIČNI PRIMJERI ZA PREPOZNAVANJE RIZIKA I PRISTUPANJE NJIHOVU RJEŠAVANJU

—

Ovo poglavlje prikazuje okvir za dubinsku analizu odgovornog poslovnog ponašanja i praktične primjere njegove provedbe za poduzeća koja sudjeluju u razvoju i upotrebi sustava umjetne inteligencije. Okvir za dubinsku analizu predstavljen u ovim Smjernicama na početku svakog koraka uključuje i pregled povezanih odredbi u postojećim okvirima, pri čemu se naznačuje na koji način svaki korak okvira za dubinsku analizu nadopunjavaju mjerodavne odredbe iz relevantnih okvira za upravljanje rizicima umjetne inteligencije i kako je s njima povezan.

U ovom se odjeljku Smjernica izlaže OECD-ov okvir za dubinsku analizu odgovornog poslovnog ponašanja za poduzeća koja sudjeluju u razvoju i upotrebi sustava umjetne inteligencije te se iznose praktični primjeri provedbe. Okvir za dubinsku analizu predstavljen u ovim Smjernicama na početku svakog koraka sadržava i pregled povezanih odredbi u postojećim okvirima, pri čemu se naznačuje na koji način svaki korak OECD-ova okvira za dubinsku analizu nadopunjavaju mjerodavne odredbe iz relevantnih okvira za upravljanje rizicima umjetne inteligencije i kako je s njima povezan.

Uz svaki korak iznose se „praktični primjeri provedbe" koji dodatno ilustriraju načine provedbe i, prema potrebi, prilagodbe popratnih mjera i postupka dubinske analize. Praktični primjeri odabrani su tako da odgovaraju kontekstu te se oslanjaju na vodeće okvire za upravljanje rizicima umjetne inteligencije, kao i na istraživanje dostupne literature te savjetovanja sa stručnjacima. Praktični primjeri nisu zamišljeni kao iscrpan popis. Neće svaki praktični primjer biti primjeren za svaku situaciju (npr. Međunarodni kodeks ponašanja u okviru Hirošimskog procesa usmjeren je na napredne sustave umjetne inteligencije). Isto tako, u nekim situacijama poduzećima mogu biti korisni dodatni primjeri ili provedbene mjere.

Tablice prikazane uz svaki korak pokazuju gdje se slični zahtjevi ili očekivanja mogu pronaći u drugim nacionalnim i međunarodnim okvirima za upravljanje rizicima umjetne inteligencije. Te tablice i povezani praktični primjeri imaju za cilj pomoći poduzećima u provedbi u više sustava nadležnosti. Iako se odredbe odnose na provedbu okvira za dubinsku analizu odgovornog poslovnog ponašanja, ova tablica nije okvir istovjetnosti jer se obuhvat i priroda očekivanja u drugim okvirima mogu razlikovati.

1. KORAK – UGRADITI ODGOVORNO POSLOVNO PONAŠANJE U POLITIKE I SUSTAVE UPRAVLJANJA

TABLICA 2.1. – 1. KORAK: PREGLED POVEZANIH ODREDBI U POSTOJEĆIM OKVIRIMA

Okvir za upravljanje rizicima umjetne inteligencije	**Povezane odredbe**
Vodič za upravljanje umjetnom inteligencijom i etiku umjetne inteligencije Saveza država Jugoistočne Azije **(„ASEAN-ov vodič")**	Odjeljak C.1: 1. Unutarnje strukture upravljanja i mjere te Prilog A: 2. Unutarnje strukture upravljanja i mjere
Australske smjernice za uvođenje umjetne inteligencije (Prakse provedbe)	Praksa provedbe 1, 2 i 4
Kanadski dobrovoljni kodeks ponašanja za odgovoran razvoj i upravljanje naprednim generativnim sustavima umjetne inteligencije i Vodič za provedbu za upravitelje sustava umjetne inteligencije **(„Kanadski kodeks ponašanja")**	Odgovornost
Okvirna konvencija Vijeća Europe o umjetnoj inteligenciji i ljudskim pravima, demokraciji i vladavini prava (Vijeće Europe 2024.) i pripadajuća metodologija za procjenu rizika i učinaka sustava umjetne inteligencije s gledišta ljudskih prava, demokracije i vladavine prava **(„HUDERIA")**	HUDERIA – tijek rada
Akt Europske unije o umjetnoj inteligenciji **(„Akt o UI-ju")**	Čl. 9. st. 1. – 3.: Sustav upravljanja rizikom, čl. 17. st. 1.: Sustav upravljanja kvalitetom
Akt EU-a o digitalnim uslugama **(„DSA")**	Čl. 14.: Uvjeti poslovanja i čl. 45.: Kodeksi ponašanja
Direktiva Europske unije o dužnoj pažnji za održivo poslovanje **(„CSDDD")**	Čl. 7.: Integriranje dužne pažnje u politike i sustave upravljanja rizicima poduzeća
Međunarodni kodeks ponašanja za napredne sustave umjetne inteligencije u sklopu Hirošimskog procesa i njegov Okvir za izvješćivanje **(„Kodeks ponašanja Hirošimskog procesa")**	Načelo 4. (točka 2.) i Načelo 5. (točke 1., 3. i 4.) te Načelo 7.
Standardni model procesa Instituta inženjera elektrotehnike i elektronike (IEEE) za rješavanje etičkih pitanja tijekom projektiranja sustava 7000-2021 **(„IEEE 7000")**	6.: Ključne uloge, 7.: Koncept operacija i postupak istraživanja konteksta
Međunarodna organizacija za normizaciju / Međunarodna elektrotehnička komisija: Informacijska tehnologija – Umjetna inteligencija – Smjernice za upravljanje rizikom 23894 **(„ISO/IEC 23894")**	5.1.: Općenito, 5.2.: Vodstvo i opredijeljenost, 5.3.: Ugrađivanje, 5.4.: Projektiranje
ISO/IEC 38507:2022 Informacijska tehnologija – Upravljanje informacijskom tehnologijom – Implikacije za upravljanje koje proizlaze iz upotrebe umjetne inteligencije u organizacijama **(„ISO/IEC 38507")**	4.: Implikacije za upravljanje koje proizlaze iz upotrebe umjetne inteligencije u organizacijama (uključujući razmatranja o upravljanju i odgovornosti); 6.: Politike za pristupanje pitanju upotrebe umjetne inteligencije (uključujući nadzor, odlučivanje, korištenje podataka, usklađenost te kulturu i vrijednosti).
ISO/IEC 42001 Informacijska tehnologija – Umjetna inteligencija – Sustav upravljanja **(„ISO/IEC 42001")**	Općenito razrađeno u poglavljima: 4.: Kontekst; 5.: Vodstvo; 6.: Planiranje; 7.: Podrška; 8.: Radni proces; 9.: Vrednovanje uspješnosti; 10.: Poboljšavanje. Provesti potrebne organizacijske i tehničke mjere radi pristupanja pitanju prepoznatih rizika. Detaljno razrađeno u Prilogu: A.2. (Politike umjetne inteligencije) i kontrolama A.2.2. – A.2.4. te, po pitanju upravljanja rizikom, u Prilogu A.5.2. (Procjena utjecaja umjetne inteligencije).
Japanske Smjernice za umjetnu inteligenciju u poslovanju, verzija 1.1 **(„Japanske Smjernice za UI u poslovanju")**	Dio 2. E. Uspostava upravljanja umjetnom inteligencijom, dijelovi 3., 4. i 5.; Dodatak 2. „Odjeljak 2. E. Uspostava upravljanja umjetnom inteligencijom", te dodaci 3., 4. i 5.
Korejski Osnovni zakon o umjetnoj inteligenciji	Čl. 34.

Okvir za upravljanje rizicima umjetne inteligencije	Povezane odredbe
Singapurski okvir za testiranje AI Verify	1.1.1. – 9.6.3.
Ministarstvo znanosti, inovacija i tehnologije Ujedinjene Kraljevine (DSIT), Smjernice: Uvod u osiguravanje pouzdanosti umjetne inteligencije **(„Okvir za osiguravanje pouzdanosti UI-ja, UK DSIT")**	3.2.: Osiguravanje pouzdanosti UI-ja i upravljanje UI-jem; 6.1.: Koraci za uspostavu osiguranja pouzdanosti UI-ja
Vodeća načela Ujedinjenih naroda o poslovanju i ljudskim pravima **(„Vodeća načela UN-a")**	Operativna načela 15. i 16.: Obveza politika
Nacionalni institut za standarde i tehnologiju Sjedinjenih Američkih Država, Okvir za upravljanje rizicima umjetne inteligencije **(„Okvir SAD-a za upravljanje rizicima UI-a")**	Upravljanje

Korak 1.1. – Politike odgovornog poslovnog ponašanja

Razraditi, donijeti i obznaniti skup politika koje se odnose na tematska područja odgovornog poslovnog ponašanja koji će ukazivati na posvećenost poduzeća načelima i standardima sadržanima u Načelima OECD-a za umjetnu inteligenciju i Smjernicama za multinacionalna poduzeća. Politike bi trebale uključivati planove za provedbu dubinske analize koja će biti mjerodavna za aktivnosti poduzeća i strana u poslovnim odnosima pri razvoju i upotrebi umjetne inteligencije.

Praktični primjeri provedbe (za sva poduzeća u lancu vrijednosti umjetne inteligencije, skupine 1 – 3)

1. Uz posvećenost mjerodavnim načelima i standardima odgovornog poslovnog ponašanja, obvezati se na provedbu Načela OECD-a za umjetnu inteligenciju, prema potrebi tijekom projektiranja, razvoja, uvođenja, rada i upotrebe sustava umjetne inteligencije (tj. usmjerenih na čovjeka, pravednih, transparentnih, objašnjivih, robusnih, sigurnih, zaštićenih i odgovornih).
2. Osmisliti ili preispitati i ažurirati postojeće politike odgovornog poslovnog ponašanja, među kojima i politike upravljanja rizikom, uz aktivno sudjelovanje dionika, uključujući radnike, predstavnike

radnika i sindikate, kako bi se uskladile s načelima iz mjerodavnih međunarodnih, regionalnih i nacionalnih okvira.
3. Na temelju rezultata procjene rizika (v. 2. korak) ažurirati ili jasnije definirati način na koji će poduzeće pristupiti najznačajnijim prepoznatim rizicima (v. 3. korak).
4. Utvrditi pragove tolerancije na rizik kako bi se odredile niska, srednja i visoka razina ozbiljnosti i vjerojatnosti rizika te utvrdili odgovarajući načini postupanja ili potaknula dublja dubinska analiza.
5. Prema potrebi učiniti politike upravljanja rizikom javno dostupnima (npr. na internetskim stranicama poduzeća te, ako je mjerodavno, na lokalnim jezicima područja na kojima poduzeće posluje ili održava poslovne odnose). U nekim će slučajevima poduzeća možda željeti određenim dionicima i poslovnim odnosima omogućiti pristup detaljnijim politikama i informacijama o upravljanju rizikom (v. 5. korak) ili se posebno obvezati u odnosu na određene rizike i pitanja.
6. Razmotriti politike koje su dovoljno fleksibilne i tehnološki neutralne te stoga ostavljaju prostora za budući razvoj, a istodobno dovoljno precizne da pruže smjernice operativnim timovima.
7. Razmotriti usklađivanje politika s mjerodavnim nacionalnim strategijama kako bi se osigurala usklađenost posvećenosti odgovornom poslovnom ponašanju i prioriteta zemlje domaćina.

Korak 1.2. – Sustavi unutarnjeg upravljanja

Nastojati ugraditi pitanja odgovornog poslovnog ponašanja i pouzdanu umjetnu inteligenciju u politike, nadzorna tijela, strukture, sustave, procese i timove poduzeća kako bi se provodili kao dio redovnih poslovnih procesa, uzimajući u obzir moguću neovisnost, autonomiju i pravnu strukturu nekih od tih tijela kako je predviđeno nacionalnim pravom i propisima.

Praktični primjeri provedbe (za sva poduzeća u lancu vrijednosti umjetne inteligencije, skupine 1 – 3)

1. Dodijeliti mjerodavnom višem rukovodstvu odgovornost za nadzor i dubinsku analizu umjetne inteligencije te upravi dodijeliti odgovornost za odgovorno poslovno ponašanje u području umjetne inteligencije u širem smislu.
2. Dodijeliti mjerodavnim odjelima odgovornost za provedbu pojedinih aspekata politika, pri čemu posebnu pozornost treba posvetiti osoblju čije bi radnje i odluke mogle imati najveći utjecaj na

povećanje ili smanjenje rizika od štetnih učinaka (kao što su razvojni timovi, osoblje uključeno u prikupljanje podataka, označavanje podataka i moderiranje sadržaja, projektiranje sustava i/ili nabavu usluga ili proizvoda te sustavi osmišljeni za donošenje odluka koje imaju značajne posljedice). Važno je da se uloge i odgovornosti te komunikacijski kanali povezani s upravljanjem rizikom dokumentiraju i priopće svim pojedincima i timovima u poduzeću.

3. Osmisliti nove ili prilagoditi postojeće sustave informiranja i evidentiranja u svrhu prikupljanja informacija o postupcima upravljanja rizicima umjetne inteligencije u vezi sa štetnim učincima (npr. postupci u okviru kojih mjerodavni timovi vode popis sustava umjetne inteligencije te dokumentiraju i priopćavaju rizike sustava umjetne inteligencije koje projektiraju, razvijaju, uvode, vrednuju i upotrebljavaju), kako bi se:
   a) osmislile politike i prakse prikupljanja, razmatranja, određivanja prioriteta i integriranja povratnih informacija dobivenih od drugih organizacijskih jedinica unutar poduzeća – izvan tima koji je razvio ili uveo sustav umjetne inteligencije – u pogledu potencijalnih rizika (npr. nabava, prodaja, usklađenost, kontrola izvoza, marketing i ljudski resursi)
   b) osmislili mehanizmi koji će omogućavati timu koji je razvio ili uveo sustave umjetne inteligencije da u projektiranje i provedbu sustava redovito uključuje vanjske i unutarnje povratne informacije o rizicima
   c) osmislili politika transparentnosti i postupci kako bi se i interno i eksterno jasno utvrdili i priopćili rizici koje predstavljaju sustavi umjetne inteligencije u svim poslovnim odnosima.
4. Uspostaviti komunikacijske kanale ili se koristiti postojećim komunikacijskim kanalima između mjerodavnog višeg rukovodstva i provedbenih odjela kako bi se omogućili razmjena i dokumentiranje informacija o riziku i odlučivanje u okviru upravljanja rizikom.
5. Priopćiti politike mjerodavnom osoblju organizacije (npr. tijekom uvođenja u posao ili osposobljavanja, tijekom postupka revizije projekta, osoblju zaduženom za upravljanje odnosima s kupcima, kao stalnu točku dnevnog reda sastanaka uprave itd.).
6. Osmisliti politike, postupke i programe osposobljavanja kako bi se osoblje upoznalo s njihovim dužnostima povezanima s upravljanjem rizikom i praksama upravljanja rizikom u organizaciji.
7. Poticati usklađivanje timova i poslovnih jedinica u pogledu

mjerodavnih vidova politika poduzeća koje se odnose na odgovorno poslovno ponašanje za umjetnu inteligenciju. To bi se moglo učiniti, na primjer, osnivanjem međusektorskih skupina ili odbora za razmjenu informacija i odlučivanje o rizicima te uključivanjem poslovnih jedinica koje mogu utjecati na odlučivanje o politikama odgovornog poslovnog ponašanja za umjetnu inteligenciju.
   a) Razmotriti uspostavu namjenskog odbora ili radne skupine s jasno definiranim ulogama i odgovornostima povezanima s provedbom dubinske analize umjetne inteligencije. Razmotriti imenovanje neovisnih vanjskih stručnjaka u sastav odbora u okviru širih aktivnosti uključivanja dionika.
   b) Osmisliti poticaje za osoblje i poslovne jedinice koji su usklađeni s politikama poduzeća o odgovornom poslovnom ponašanju za umjetnu inteligenciju (npr. uključivanje ciljeva ili pokazatelja u ocjenjivanje zaposlenika povezanih s provedbom politika odgovornog poslovnog ponašanja, kao što su potrošnja energije, provedba aktivnosti uključivanja dionika i donošenje politika odgovornog poslovnog ponašanja; te izazovi povezani s nagradama odnosno hakatoni usmjereni na pronalaženje rješenja za izazove u području odgovornog poslovnog ponašanja).
8. Preispitati postojeće procese u području informacijske tehnologije, sigurnosti, nabave, životnog ciklusa razvoja softvera (SDLC) itd. kako bi se utvrdilo na koji se način mogu uskladiti s politikama i procesima uspostavljenima u vezi s dubinskom analizom umjetne inteligencije.
9. Poticati široko sudjelovanje u odlučivanju povezanom s upravljanjem rizicima umjetne inteligencije (osigurati, primjerice, da u odlučivanju sudjeluju mjerodavno osoblje ili dionici različitih struka, s različitim iskustvom, stručnim znanjem i pozadinama).
10. Razviti sustave za praćenje incidenata i odgovora na njih.
11. Osmisliti nove ili iskoristiti odnosno prilagoditi postojeće postupke podnošenja pritužbi ili zviždanja kako bi osoblje moglo upozoravati na sporna pitanja ili podnositi pritužbe u vezi s tematskim područjima odgovornog poslovnog ponašanja u skladu s nacionalnim propisima.
12. Osmisliti politike i prakse za poticanje kritičkog razmišljanja pri projektiranju, razvoju, uvođenju i upotrebi sustava umjetne inteligencije.
13. Osmisliti sigurne procese nadogradnje, stavljanja izvan upotrebe i postupnog ukidanja sustava umjetne inteligencije kojima se neće

povećati postojeći rizici, stvoriti novi rizici niti smanjiti pouzdanost poduzeća.

14. Izraditi planove za izvanredne situacije radi postupanja u slučaju kvarova, incidenata ili štetnih učinaka povezanih sa sustavima umjetne inteligencije.
15. Izraditi plan uključivanja dionika kako bi im se omogućilo da ocjenjuju i prate provedbu mjerodavnih procesa odgovornog poslovnog ponašanja u svim dijelovima poduzeća, osiguravajući da u tim procesima redovito sudjeluju dionici, uključujući radnike, predstavnike radnika i sindikate.
16. Uspostaviti ili se uključiti u zajedničke inicijative za razvoj, unapređivanje i, prema potrebi, donošenje zajedničkih standarda, alata, mehanizama i najboljih praksi za osiguravanje sigurnosti, zaštite i pouzdanosti naprednih sustava umjetne inteligencije, kao što je OECD-ov Katalog alata za pouzdanu umjetnu inteligenciju.

Korak 1.3. – Očekivanja od strana u poslovnim odnosima

Uključiti očekivanja u pogledu odgovornog poslovnog ponašanja i politike odgovornog poslovnog ponašanja u suradnju sa stranama u poslovnim odnosima.

Praktični primjeri provedbe (za sva poduzeća u lancu vrijednosti umjetne inteligencije, skupine 1 – 3)

1. Priopćavati aspekte politika odgovornog poslovnog ponašanja ključne za umjetnu inteligenciju mjerodavnim stranama u poslovnim odnosima, uključujući dobavljače ulaznih podataka, prodajne partnere i korisnike sustava umjetne inteligencije.
2. Ako se poduzeća oslanjaju na prodajne partnere – strane u poslovnim odnosima koje kupuju, distribuiraju, integriraju i preprodaju proizvode i usluge krajnjim korisnicima – uspostaviti komunikacijske kanale duž prodajnog lanca i s mjerodavnim vanjskim dionicima kako bi se osigurala kontinuirana dubinska analiza.
3. Osmišljavati i provoditi pretkvalifikacijske postupke za dobavljače ili kupce pri kojima se uzima u obzir dubinska analiza mjerodavnih tematskih područja odgovornog poslovnog ponašanja, po mogućnosti prilagođavajući te postupke specifičnom riziku i kontekstu, usredotočivši se na ona tematska područja koja su prepoznata kao mjerodavna za strane u poslovnim odnosima i njihove aktivnosti ili područja poslovanja.
4. Prema potrebi, osobito u pogledu poslovnih odnosa s MSP-ovima, razmotriti da se dobavljačima, kupcima, krajnjim korisnicima i drugim stranama u poslovnim odnosima osiguraju odgovarajući resursi kako bi razumjeli i primjenjivali mjerodavne politike odgovornog poslovnog ponašanja te provodili dubinsku analizu (npr. uključivanje mjerodavnih strana u poslovnim odnosima i radnika u ciljane aktivnosti informiranja, osposobljavanja ili izgradnje kapaciteta). U idealnim bi okolnostima, kad god je to moguće, resursi i dodatne smjernice koji se osiguravaju kupcima i krajnjim korisnicima trebali biti maksimalno konkretni i ciljani.
5. Nastojati razumjeti i ukloniti prepreke koje proizlaze iz načina poslovanja poduzeća koje bi mogle ometati sposobnost dobavljača, korisnika i drugih strana u poslovnim odnosima da provode politike odgovornog poslovnog ponašanja, kao što su prakse nabave koje provodi poduzeće pri nabavi algoritama, skupova podataka, softvera ili hardvera.

2. KORAK – PREPOZNATI I PROCIJENITI STVARNE I POTENCIJALNE ŠTETNE UČINKE

TABLICA 2.2. – 2. KORAK: PREGLED POVEZANIH ODREDBI U POSTOJEĆIM OKVIRIMA

ASEAN-ov vodič	Odjeljak C.2: Utvrđivanje razine uključenosti ljudi u odlučivanje potpomognuto umjetnom inteligencijom; C.3: Upravljanje operacijama; i Prilog A: 3.: Utvrđivanje razine uključenosti ljudi u odlučivanje potpomognuto umjetnom inteligencijom
Australske smjernice za uvođenje umjetne inteligencije (Prakse provedbe)	Praksa provedbe 2. i 3.2.
Kanadski kodeks ponašanja	1. sigurnosna mjera
HUDERIA	Analiza rizika utemeljena na kontekstu, procjena učinka

Akt EU-a o UI-ju	Čl. 9. st. 2.: Utvrđivanje, analiza i evaluacija poznatih i predvidljivih rizika, čl. 55. st. 1. Obveze za UI modele opće namjene sa sistemskim rizikom
DSA	Čl. 34.: Procjena rizika
CSDDD	Čl. 8. i 9.: Utvrđivanje i procjena stvarnih i potencijalnih negativnih učinaka te, prema potrebi, rangiranje potencijalnih i stvarnih negativnih učinaka po prioritetu
Kodeks ponašanja Hirošimskog procesa	Načelo 1., 2., 6. i 7.
IEEE 7000	8. Proces utvrđivanja i određivanja prioriteta etičkih vrijednosti; 9. Proces definiranja etičkih zahtjeva
ISO 31000 i ISO/IEC 23894	6.3.: Područje primjene, kontekst, kriteriji – 6.4.: Procjena rizika
ISO/IEC 42001	6.: Planiranje (6.1.1., 6.1.2., 6.1.4.); 8.: Radni proces (8.2., 8.4.); Prilog A.5. (Procjena utjecaja sustava umjetne inteligencije); kontrole A.5.2. – A.5.5.
ISO/IEC 42005	5.8. Stvarni i potencijalni utjecaji
Japanske Smjernice za umjetnu inteligenciju u poslovanju	Prilog 1.B. Koristi i rizici umjetne inteligencije; Prilog 2.A. Uspostava upravljanja umjetnom inteligencijom i njezina praćenja na razini menadžmenta
Korejski Osnovni zakon o umjetnoj inteligenciji	Čl. 32., 33. i 35.
Singapurski okvir za testiranje AI Verify	Sigurnost 4.1.1. – 4.3.1.
Okvir za osiguravanje pouzdanosti UI-ja, UK DSIT	4.1.1.: Mjerenje; 4.1.2.: Vrednovanje; 5.4.: Procjena rizika; 5.5.: Procjena učinka; 5.6.: Provjera nepristranosti
Vodeća načela UN-a o poslovanju i ljudskim pravima	Operativna načela 17. i 18.
Okvir SAD-a za upravljanje rizicima UI-a	Upravljanje 1, 4, Mapiranje 1 – 5

Korak 2.1. – Početno određivanje obuhvata rizika

Odrediti obuhvat kako bi se utvrdilo gdje se rizici mogu pojaviti i gdje mogu biti najizraženiji.

Na međunarodnoj, regionalnoj i nacionalnoj razini postoje brojni okviri koji opisuju rizike povezane s razvojem i upotrebom sustava umjetne

inteligencije te preporučuju mjere koje bi poduzeća trebala poduzeti radi pristupanja problemu tih rizika. Jedan od ciljeva ovih Smjernica jest poduprijeti provedbu drugih okvira za upravljanje rizikom u poduzećima. Iako popis okvira može poslužiti kao polazna točka za prepoznavanje niza potencijalno mjerodavnih rizika u okviru dubinske analize pojedinog poduzeća, on nije iscrpan, a mnogi se rizici preklapaju i mogu biti međusobno povezani. Isto tako, nisu svi okviri mjerodavni za svako poduzeće. Od svakog se poduzeća očekuje da utvrdi prioritetna područja rizika na temelju okolnosti u kojima posluje.

Pri određivanju redoslijeda pristupanja rizicima (v. korak 2.4.) poduzeća bi trebala uzeti u obzir da su određeni rizici tijesno povezani s drugima ili mogu omogućiti njihov nastanak.

Baš kao što je slučaj s mnogim novim tehnologijama, zlonamjerni akteri iz javnog i privatnog sektora mogu pronaći načine za zloupotrebu sustava umjetne inteligencije. Značajan dvojaki potencijal sustavâ umjetne inteligencije i mogućnosti njihove prenamjene mogu voditi štetnoj upotrebi čak i ako je namjena za koju su projektirani bezazlena.

Praktični primjeri provedbe (za dobavljače ulaznih elemenata za umjetnu inteligenciju, skupina 1)

1. Prepoznati strane u poslovnim odnosima koje aktivno sudjeluju u razvoju sustava umjetne inteligencije (tj. strane u poslovnim odnosima u skupini 2).
2. Razviti početno razumijevanje vrsta sustava umjetne inteligencije koje razvijaju strane u poslovnim odnosima, uključujući informacije o rizicima povezanima sa sustavima umjetne inteligencije opisane u koraku 2.1.

Praktični primjeri provedbe (za poduzeća uključena u životni ciklus sustava umjetne inteligencije, skupina 2)

1. Uspostaviti razumijevanje o ulozi poduzeća tijekom životnog ciklusa sustava umjetne inteligencije i voditi ažuran registar sustava umjetne inteligencije povezanih s poduzećem.
2. Uspostaviti razumijevanje o rizicima potencijalnih štetnih učinaka povezanih s razvojem i/ili upotrebom sustava umjetne inteligencije. To je moguće postići analizom dostupnih izvora i savjetovanjem o tome koji bi rizici mogli biti povezani sa sustavom umjetne inteligencije te prikupljanjem i pregledavanjem izvješća o rizicima koji se odnose na sustav umjetne inteligencije ili poduzeće koje

je sustav razvilo ili izmijenilo. Interni izvori informacija o rizicima uključuju mehanizme praćenja incidenata, tijela za nadzor i komunikacijske kanale opisane u koraku 1.2. Vanjski izvori informacija o rizicima uključuju izvješća nacionalnih institucija za ljudska prava, nacionalnih opservatorija za umjetnu inteligenciju, regulatornih tijela, resornih ministarstava, međunarodnih i regionalnih mehanizama odgovornosti za ljudska prava, organizacija civilnog društva te radnika, predstavnika radnika, sindikata, javnih baza podataka o incidentima,[11] sudskih predmeta, mehanizama za podnošenje pritužbi te uključivanje pogođenih zajednica, kao i onih izloženih riziku. Okvir OECD-a za klasifikaciju sustava umjetne inteligencije također pruža temelj na kojem se može graditi razumijevanje rizika povezanih sa sustavima umjetne inteligencije (v. Okvir OECD-a za klasifikaciju sustava umjetne inteligencije (OECD, 2022[13])).

3. Informacije o rizicima mogu obuhvaćati sljedeća područja:
   a) *interakcija sustavâ umjetne inteligencije.* Iako dva sustava umjetne inteligencije mogu biti bezopasna kad se promatraju odvojeno, njihova sposobnost nanošenja štete može se povećati ako se uvedu na način koji omogućuje njihovu međusobnu interakciju bez zaštitnih mehanizama
   b) priroda sustava umjetne inteligencije, slučaj upotrebe ili proizvod u kojem se koristi (npr. tehnologija prepoznavanja lica i emocija, prediktivno policijsko djelovanje, tehnologija nadzora, marketinška tehnologija)
   c) vrsta korisnika / način upotrebe sustava umjetne inteligencije i poslovni ciljevi korisnika (npr. medijske kuće, pružatelji zdravstvenih usluga, pravosudna tijela, tijela za provedbu zakona, obavještajne agencije, pojedinci)
   d) izvori ulaznih podataka, softver, fizičke komponente i aspekti sustava koji uključuju čovjeka (npr. rizici za radnike tijekom označavanja podataka i moderiranja sadržaja, pribavljanje osobnih podataka ili intelektualnog vlasništva)
   e) geografski/socioekonomski/politički kontekst u kojem se uvodi sustav umjetne inteligencije (npr. područja s visokom razinom korupcije, kršenjima ljudskih i radnih prava te područja sukoba)
   f) kompetentnost i znanstvena valjanost sustava umjetne inteligencije (npr. rizici od nekompetentnog ili neprimjerenog obavljanja važnih zadaća)
   g) poznate ili razumno predvidive okolnosti povezane s upotrebom sustava umjetne inteligencije u skladu s njegovom namjenom ili u uvjetima razumno predvidive nepravilne upotrebe ili zloupotrebe, kad takve okolnosti mogu imati za posljedicu štetne učinke (za pokazatelje upotrebe sustava umjetne inteligencije koji mogu predstavljati veći rizik od štetnih učinaka v. okvir 2.1.).
4. Ako nije izvedivo ili praktično provoditi detaljne procjene svih sustava umjetne inteligencije, razmotriti uvođenje sustava eskalacije (npr. uvođenjem upitnika za sve nove sustave umjetne inteligencije radi procjene osnovne razine rizika, pri čemu se svi sustavi kod kojih postoje pokazatelji visokog rizika upućuju na daljnju, detaljnu dubinsku analizu).
5. Redovito preispitivati rezultate određivanja obuhvata te pri nastanku značajnih promjena (npr. poslovanje u novoj zemlji, razvoj novog sustava ili proizvoda umjetne inteligencije, restrukturiranje, uspostava novih poslovnih odnosa, značajne promjene mjerodavnog zakonodavstva).
6. Uzimati u obzir mjerodavne zakone, propise i norme u područjima kao što su zaštita potrošača ili sektorski propisi, zakoni i norme (npr. u zdravstvu, proizvodnji, zrakoplovstvu itd.) kako bi se poduzećima pomoglo u razumijevanju rizika i utvrđivanju visokorizičnih načina upotrebe sustava umjetne inteligencije.

OKVIR 2.1. PRIMJERI VISOKORIZIČNIH NAČINA UPOTREBE SUSTAVA UMJETNE INTELIGENCIJE PREUZETI IZ RAZLIČITIH OKVIRA

Visokorizični načini upotrebe sustava umjetne inteligencije ovise o kontekstu, a različiti sustavi nadležnosti imaju različite kriterije za utvrđivanje tih načina. U nekima od vodećih okvira za upravljanje rizicima umjetne inteligencije su kao potencijalno visokorizične prepoznate sljedeće primjene umjetne inteligencije:

- načini upotrebe koji predstavljaju kemijske, biološke, radiološke i nuklearne rizike, primjerice mogućnosti smanjenja prepreka za pristup oružju, uključujući njegov razvoj, projektiranje ili nabavu. U tom je kontekstu važno uzeti u obzir i aspekte dvojne namjene
- načini upotrebe u kritičnoj infrastrukturi koji mogu predstavljati rizike za zdravlje i sigurnost (npr. promet)
- načini upotrebe u obrazovanju ili strukovnom osposobljavanju koji mogu odrediti ili snažno utjecati na pristup obrazovanju i profesionalni tijek nečijeg života (npr. ocjenjivanje ispita, sustavi dodjele)
- načini upotrebe povezani sa savjetovanjem u području mentalnog zdravlja i osiguravanjem društva putem simulirane interakcije, osobito kad uključuju maloljetnike ili druge ranjive pojedince
- načini upotrebe u zapošljavanju, upravljanju radnicima i pristupu zapošljavanju (npr. softver za razvrstavanje životopisa u postupcima zapošljavanja)
- načini upotrebe povezani s odobravanjem ili uskraćivanjem pristupa osnovnim privatnim i javnim uslugama (npr. pristup zdravstvenoj skrbi ili bodovanje kreditne sposobnosti koje utječe na mogućnost građana da dobiju kredit)
- načini upotrebe povezani s provedbom zakona i kaznenim pravosuđem (npr. informiranje o odmjeravanju kazne, uvjetnom otpustu i probaciji, odluke o puštanju na slobodu prije suđenja ili zadržavanju u pritvoru, nadzor, predviđanje kaznenih djela i prediktivno policijsko djelovanje te forenzička analiza)
- načini upotrebe povezani s upravljanjem migracijama, azilom i nadzorom granica (npr. provjera vjerodostojnosti putnih isprava ili odlučivanje o zahtjevima za azil).

Izvori: čl. 6. Akta o umjetnoj inteligenciji te prilozi I. i III. Uredbi (EU) 2024/1689 Europskog parlamenta i Vijeća od 13. lipnja 2024. o utvrđivanju usklađenih pravila o umjetnoj inteligenciji i o izmjeni uredaba (EZ) br. 300/2008, (EU) br. 167/2013, (EU) br. 168/2013, (EU) 2018/858, (EU) 2018/1139 i (EU) 2019/2144 te direktiva 2014/90/EU, (EU) 2016/797 i (EU) 2020/1828 (Europska unija, 2024[14]), https://eur-lex.europa.eu/legal-content/HR/ALL/?uri=CELEX:32024R1689; G7 (2023[15]) Međunarodni kodeks ponašanja u okviru Hirošimskog procesa za napredne sustave umjetne inteligencije, https://www.mofa.go.jp/ecm/ec/page5e_000076.html.

Praktični primjeri provedbe (za korisnike sustava umjetne inteligencije, skupina 3)

1. Uspostaviti početno razumijevanje svih načina upotrebe sustava umjetne inteligencije u poduzeću, uključujući informacije o rizicima povezane sa sustavima umjetne inteligencije opisane pod korakom 2.1., točka 54. (b) (npr. ljudski resursi, marketing, prodaja, služba za korisnike, nabava, dubinska analiza itd.) te razmotriti koji načini upotrebe zahtijevaju dublju dubinsku analizu. U slučajevima niskog rizika možda neće biti opravdane daljnje detaljne aktivnosti dubinske analize.
2. Prepoznati strane u poslovnim odnosima koje razvijaju i uvode sustave umjetne inteligencije koji se koriste u poslovanju, proizvodima i uslugama.

Korak 2.2. – Detaljna procjena najznačajnijih rizika

Počevši od onih područja rizika koja su prepoznata kao najznačajnija, opetovano provoditi sve detaljnije procjene prioritetnih rizika povezanih 1.) s aktivnostima poduzeća i 2.) sa stranama u poslovnim odnosima s poduzećem (npr. dobavljači, kupci i korisnici).

OKVIR 2.2. RAZUMIJEVANJE PRISTUPA UTEMELJENOG NA RIZIKU

U Smjernicama za multinacionalna poduzeća poduzećima se preporučuje pristup štetnim učincima utemeljen na riziku, uviđajući da možda neće uvijek biti moguće odmah odgovoriti na sve štetne učinke. Taj je koncept ključan za primjenu dubinske analize na umjetnu inteligenciju, osobito sustave umjetne inteligencije opće namjene, pri čemu bi neka poduzeća mogla biti povezana s učincima koje izaziva više stotina ili tisuća strana u poslovnim odnosima koje upotrebljavaju sustav umjetne inteligencije.

U tom pogledu se u Smjernicama za multinacionalna poduzeća pojašnjava kako se poduzeća, ako imaju velik broj strana u poslovnim odnosima, potiče da prepoznaju opća područja u kojima je rizik od štetnih učinaka najveći te da u skladu s time na temelju te procjene rizika odrede prioritete za dubinsku analizu. Pristup utemeljen na riziku bi trebao uzeti u obzir i poznate ili razumno predvidive okolnosti povezane s upotrebom sustava umjetne inteligencije u skladu s njegovom namjenom ili u uvjetima njegove razumno predvidive nepravilne upotrebe ili zloupotrebe.

Na značaj (ili važnost) štetnog učinka ukazuje stupanj njegove vjerojatnosti i ozbiljnosti (OECD, 2018[3]). Ozbiljnost učinaka prosuđuje se na temelju njihova razmjera, obuhvata i neotklonjive naravi (v. tablicu 2.3.). Razmjer se odnosi na težinu štetnog učinka. Obuhvat se odnosi na domašaj učinka, na primjer na broj pojedinaca koji jesu ili će biti pogođeni. Neotklonjiva narav znači bilo kakvo ograničenje sposobnosti povratka na situaciju jednaku onoj prije štetnog učinka. Ozbiljnost nije apsolutni pojam, već je specifična za kontekst. Pojam vjerojatnosti povezan je s čimbenicima ozbiljnosti te bi ga trebalo razmatrati zajedno s njima.

Na primjer, ako sustav umjetne inteligencije predstavlja rizike koji su manje vjerojatni, ali vrlo ozbiljni, kao i rizike koji su vrlo vjerojatni, ali manje ozbiljni, od poduzeća uključenih u razvoj sustava umjetne inteligencije očekuje se da pokažu da se oba rizika uzimaju u obzir i prate te da su uspostavljeni postupci odgovora na rizik koji poduzeće u tom trenutku smatra najznačajnijim.

Od poduzeća se očekuje da imaju uspostavljen vjerodostojan postupak određivanja prioriteta te da napreduju u ostvarivanju ciljeva usmjerenih na rezultate i s jasno definiranim rokovima. Uključivanje dionika u određivanje prioriteta rizika te transparentni kriteriji za određivanje prioriteta, kao i njihova obrazloženja, doprinijet će jačanju vjerodostojnosti postupaka dubinske analize (v. i okvir 2.6.).

Praktični primjeri provedbe (za sva poduzeća uključena u životni ciklus sustava umjetne inteligencije, skupina 2) – prepoznavanje rizika u svojem poslovanju

1. Izraditi popis konkretnih pravnih zahtjeva i nacionalnih/međunarodnih/strukovnih standarda primjenjivih na sustav ili subjekt u području umjetne inteligencije koji predstavlja predmet procjene, uključujući mjerodavne nacionalne i međunarodne standarde odgovornog poslovnog ponašanja i rada.
2. Pregledati informacije o testiranju, evaluaciji, verifikaciji i validaciji (TEVV), uključujući one povezane s eksperimentalnim dizajnom, prikupljanjem i odabirom podataka (npr. dostupnost, točnost, reprezentativnost, prikladnost), pouzdanošću sustava i validacijom konstrukta. Nastojati osigurati da alati ili pokazatelji koji se koriste za mjerenje, testiranje ili ublažavanje rizika sustava umjetne inteligencije i sami budu testirani, da imaju dokazanu i mjerljivu upotrebnu vrijednost te da su prošli testiranja u pogledu osiguranja kvalitete za sve pokazatelje.
3. Prema potrebi, s obzirom na rizik, pregledati evaluacije sustava umjetne inteligencije koje uključuju ljudske ispitanike.
4. Preispitati način na koji se očekuje da će ljudi upotrebljavati i nadzirati izlazne podatke sustava.
5. Savjetovati se sa stručnjacima za predmetno područje, korisnicima i poduzećima izvan tima koji je razvio ili uveo sustav umjetne inteligencije.
6. Savjetovati se i uključiti dionike, uključujući radnike, predstavnike radnika i sindikate, zajednice koje jesu ili bi mogle postati izložene štetnim učincima, neovisne stručnjake i organizacije civilnog društva radi prikupljanja informacija o značajnim rizicima, uzimajući u obzir moguće prepreke učinkovitoj uključenosti dionika. Ako izravno savjetovanje s dionicima nije moguće, razmotriti razumne alternative kao što su savjetovanje s neovisnim stručnjacima, uključujući branitelje ljudskih prava, sindikate i skupine civilnog društva.
   a) Savjetovati se s dionicima izloženima mogućim štetnim učincima prije i tijekom projekata ili aktivnosti koje bi mogle utjecati na njih.
   b) Razmotriti mogućnost da se nekim dionicima omogući sudjelovanje u testiranjima sustava umjetne inteligencije ili njihovo preispitivanje (npr. A/B testovi), uz dužno uvažavanje zaštite povjerljivih poslovnih informacija i prava intelektualnog vlasništva.
7. Razmotriti rizike od štetnih učinaka u fazi prije uvođenja i u fazi razvoja (npr. krađa modela i zloupotreba u okviru interne upotrebe).
8. Prepoznati rizike za robusnost i sigurnost sustava umjetne inteligencije, primjerice putem matematičkih jamstava ili testiranja otpornosti na protivničke napade.
9. Prepoznati rizike za privatnost i upravljanje podacima na razini podataka i modela (v. okvir 2.3.).
10. Utvrditi rizike od toga da sustav umjetne inteligencije omogućava ili zagovara ishode koji izazivaju štetne učinke na ljudska prava te nanose štetu društvu i javnom interesu.

OKVIR 2.3. PREPOZNAVANJE RIZIKA ZA KVALITETU PODATAKA, INTEROPERABILNOST I PRISTUP TIJEKOM ŽIVOTNOG CIKLUSA SUSTAVA UMJETNE INTELIGENCIJE

Prikupljanje i obrada podataka u kontekstu umjetne inteligencije nose niz rizika koji, ako se ne riješe na odgovarajući način, mogu narušiti vjerodostojnost sustava umjetne inteligencije i točnost njegovih izlaznih podataka te imati za posljedicu štetne učinke. Učinci mogu biti posljedica nezakonito ili neprimjereno pribavljenih podataka, manipuliranih podataka i asimetričnog pristupa podacima. Mogu nastati na razini podataka i modela, na njihovu sjecištu, kao i tijekom interakcije čovjeka i sustava umjetne inteligencije.

- Na razini podataka: procjene učinka na zaštitu podataka predstavljaju standardni postupak procjene rizika. Taj je postupak u nekim sustavima nadležnosti pravno formaliziran. U okviru tih procjena uzimaju se u obzir rizici trovanja podataka, pri čemu se zlonamjerno manipulira podacima za učenje kako bi se utjecalo na ponašanje modela.
- Na razini modela: sigurnost modela umjetne inteligencije može se procjenjivati na temelju:
  1. razine pristupa koju bi mogao imati zlonamjerni akter, od „crne kutije" (npr. ne poznajući model) do „potpune transparentnosti" (npr. raspolažući potpunim informacijama o modelu i njegovim podacima za učenje)
  2. faza u kojima može doći do napada (npr. tijekom učenja ili u operativnoj fazi)
  3. vjerojatnosti pasivnih (npr. „pošten, ali znatiželjan") ili aktivnih (npr. potpuno zlonamjernih) napada s obzirom na profil prijetnje
  4. toga oslanja li se model na prekogranične prijenose podataka (npr. ako je pružatelj usluga razvojnog inženjeringa multinacionalno poduzeće i podaci se prikupljaju iz više sustava nadležnosti ili ako se dio razvoja modela povjerava drugim vanjskim subjektima).

- Na sjecištu razine podataka i modela: rizici uključuju izvođenje zaključaka o određenim članovima skupa podataka za učenje putem interakcije s modelom. Tehnike za procjenu razina ranjivosti uključuju statističko otkrivanje podataka, inverziju modela, izvođenje zaključaka o predstavnicima razreda te zaključivanje o članstvu i svojstvima.
- Na razini interakcije čovjeka i umjetne inteligencije: osposobljavanje, kontrolni popisi i postupci verifikacije mogli bi pomoći u prepoznavanju rizika koji proizlaze iz interakcije čovjeka i sustava (npr. nenamjerne radnje – ili izostanak radnji – razvojnih inženjera ili korisnika koji ugrožavaju privatnost ili upravljanje podacima sustava umjetne inteligencije).

Smjernice OECD-a o zaštiti privatnosti, donesene 1980. i revidirane 2013. (OECD, 2015[16]), temelj su rada OECD-a u području zaštite privatnosti i prepoznate su kao globalni minimalni standard za zaštitu privatnosti i osobnih podataka. Osim toga, osobito poglavlje o odgovornosti Smjernica za provedbu Smjernica OECD-a o zaštiti privatnosti može pomoći dionicima da bolje razumiju i primjenjuju načelo odgovornosti utvrđeno u Smjernicama o zaštiti privatnosti kroz programe upravljanja privatnošću. Takvi programi i njihov pristup utemeljen na riziku osiguravaju organizacijama vrijedan alat za suočavanje s razvojnim rizicima i izazovima, kao što su oni koje donose nove tehnologije.

Izvor: OECD (2023[17]), *OECD Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI*, https://doi.org/10.1787/2448f04b-en; OECD (2013[18]), *Guidelines governing the protection of privacy and transborder flows of personal data*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188.

Praktični primjeri provedbe (za poduzeća uključena u životni ciklus sustava umjetne inteligencije, skupina 2) – prepoznavanje rizika povezanih sa stranama u poslovnim odnosima koje upotrebljavaju sustave umjetne inteligencije

1. Mapirati mjerodavne aktivnosti organizacije ili strane u poslovnim odnosima povezane s rizikom koji je utvrđen kao prioritetan.
2. Preispitati imaju li mjerodavne strane u poslovnim odnosima visokog rizika uspostavljene politike dubinske analize i interne sustave upravljanja (u skladu s 1. korakom).
3. Pri procjeni strana u poslovnim odnosima, osobito pružatelja usluga obogaćivanja podataka, prema dostupnosti koristiti informacije iz procjena učinaka same organizacije ili trećih strana, pravnih

pregleda, sustava upravljanja usklađenošću, financijskih revizija, inspekcija zaštite na radu te zdravlja i sigurnosti na radnom mjestu, izvješća organizacija radnika, sindikata i/ili organizacija civilnog društva te svih drugih mjerodavnih procjena koje provode organizacija ili strukovne inicijative odnosno inicijative koje uključuju više dionika, kao i postupaka „upoznaj svojeg klijenta" (KYC).

4. Razmotriti uvođenje sustava eskalacije za označavanje potencijalno visokorizičnih prodaja bez nepotrebnog opterećivanja prodaja niskog rizika. U nekim kontekstima taj postupak može biti povezan s postojećim postupcima usklađenosti za kontrolu izvoza i sankcije ili se može integrirati u njih. To bi moglo uključivati:
   a) utvrđivanje kriterija za označavanje visokorizičnih prodaja i provedbu dodatne dubinske analize prije odobravanja takvih prodaja
   b) osposobljavanje mjerodavnog osoblja na temu rizika krajnje upotrebe i dubinske analize klijenata
   c) ugrađivanje politika odgovornog poslovnog ponašanja u poticaje za uspješnost prodajnih timova i prodajnih partnera
   d) ugrađivanje zahtjeva odgovornog poslovnog ponašanja u ugovore s prodajnim partnerima
   e) osiguravanje osposobljavanja i smjernica prodajnim partnerima o dubinskoj analizi odgovornog poslovnog ponašanja
   f) uspostavu mehanizma izvješćivanja o rizicima za prodajne partnere.

Praktični primjeri provedbe (za dobavljače ulaznih elemenata i korisnike sustava umjetne inteligencije, skupine 1 i 3)

1. Pri suradnji s poduzećima uključenima u aktivnosti skupine 2 koja su u postupcima dubinske analize označena kao visokorizična, poduzeća bi se trebala uključiti izravno u komunikaciju sa stranama u poslovnim odnosima kako bi se bolje upoznala s njihovim aktivnostima dubinske analize te ih, prema potrebi, potaknula da poduzmu dodatne korake kako bi pristupila pitanju stvarnih i potencijalnih štetnih učinaka. Neke od konkretnih informacija u okviru dubinske analize koje treba prikupiti od strana u poslovnim odnosima jesu:
   a) obveze koje je akter u području umjetne inteligencije preuzeo u pogledu poštovanja nacionalnih ili međunarodnih standarda za odgovornu umjetnu inteligenciju (npr. Kodeks ponašanja u okviru Hirošimskog procesa, Pakt EU-a o UI-ju, mjerodavni kodeksi ponašanja proizašli iz Akta EU-a o UI-ju i/ili Akta o digitalnim uslugama (DSA))
   b) značajni štetni učinci ili rizici sustava umjetne inteligencije koji su prepoznati, određeni kao prioritetni i procijenjeni
   c) kriteriji za određivanje prioriteta rizika koje treba rješavati (v. korak 2.4.)
   d) radnje koje su poduzete ili planirane radi sprečavanja ili ublažavanja rizika, gdje je to moguće, uključujući procijenjene rokove i referentne vrijednosti za poboljšanje i njihove ishode (v. 3. korak)
   e) mjere za praćenje provedbe i rezultata (v. 4. korak)
   f) osiguravanje otklanjanja štete ili suradnja na njezinu otklanjanju (v. 6. korak).
2. Ako informacije u okviru dubinske analize nisu dostupne ili strane u poslovnim odnosima ne osiguraju dovoljno detaljne podatke za potrebe procjene rizika poduzeća, poduzeća se mogu koristiti postojećim procjenama – kao što su procjene koje se zajednički koriste u okviru stručnih suradnji ili inicijativa koje uključuju više dionika, akademskih studija ili pružatelja usluga ocjenjivanja sigurnosti umjetne inteligencije te izvješća nevladinih organizacija ili drugih pružatelja usluga istraživanja tržišta – uz nastavak suradnje sa stranama u poslovnim odnosima kako bi se osigurala dostupnost tih informacija.

Korak 2.3. – Procijeniti uključenost u stvarni ili potencijalni štetni učinak (uzrok, doprinos, izravna povezanost)

Procijeniti uključenost poduzeća u prepoznate stvarne ili potencijalne štetne učinke. Konkretno, procijeniti je li organizacija ili mjerodavna strana u poslovnom odnosu prouzročila (ili bi mogla prouzročiti) štetan učinak ili doprinijela (ili bi mogla doprinijeti) štetnom učinku odnosno je li štetan učinak izravno povezan s njihovim poslovanjem, proizvodima ili uslugama (ili bi to mogao postati). Odnos poduzeća prema štetnom učinku nije statičan. Može se promijeniti, na primjer ovisno o razvoju situacija i stupnju do kojeg se, dubinskom analizom i koracima poduzetima kako bi se pristupilo rješavanju utvrđenih rizika i štetnih učinaka, smanjio rizik od nastanka učinaka.

OKVIR 2.4. RAZUMIJEVANJE UKLJUČENOSTI U RIZIK

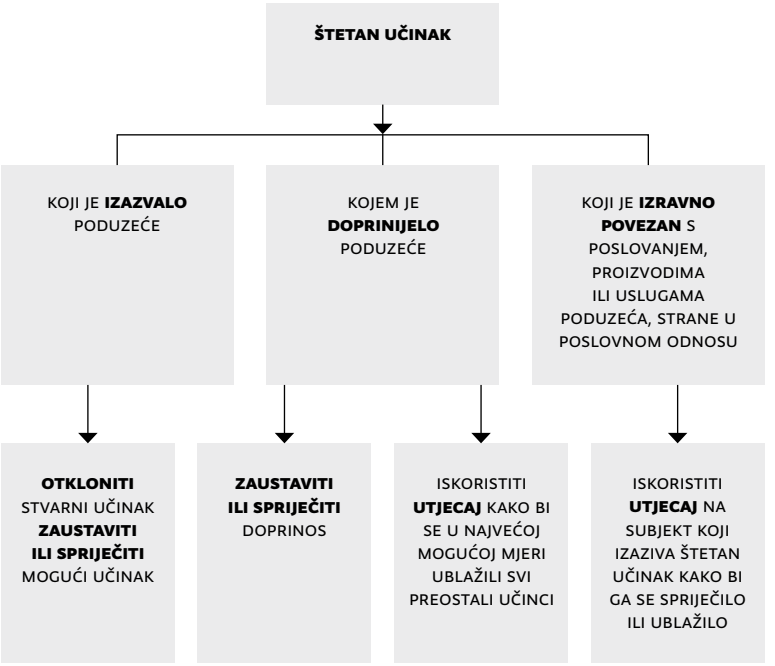
Iako se od svih poduzeća očekuje provođenje dubinske analize, razina dubinske analize razlikuje se ovisno o uključenosti u stvarni ili potencijalni štetni učinak. Od poduzeća koja izazivaju štetne učinke očekuje se da zaustave ili spriječe potencijalne učinke te da otklone štetu uzrokovanu stvarnim učincima.

Od poduzeća koja doprinose štetnim učincima očekuje se da zaustave ili spriječe svoj doprinos potencijalnim učincima, da otklone svoj doprinos šteti te da iskoriste, a prema potrebi i uspostave utjecaj na strane u poslovnim odnosima kako bi spriječila ili ublažila dodatne rizike. Poduzeće doprinosi učinku ako ga izazivaju njegove aktivnosti u kombinaciji s aktivnostima drugih subjekata ili ako aktivnosti poduzeća imaju za posljedicu da neki drugi subjekt izazove štetan učinak ili mu to olakšavaju ili ga potiču na to. Doprinos mora biti znatan što znači da se ne razmatraju manji ili trivijalni doprinosi.

Od poduzeća koja su izravno povezana sa štetnim učincima očekuje se da iskoriste, a prema potrebi i uspostave utjecaj kako bi potaknula strane u poslovnim odnosima na sprečavanje ili ublažavanje štetnih učinaka ili rizika. „Povezanost" je definirana odnosom između štetnog učinka i proizvoda, usluga ili operacija poduzeća posredstvom drugog subjekta (odnosno strane u poslovnim odnosima). Doprinos ili izravna povezanost sa štetnim učinkom nakon što je sustav umjetne inteligencije uveden, prodan ili preprodan često se mogu povezati s neublaženim rizicima u dizajnu proizvoda ili visokorizičnim krajnjim korisnicima.

Za primjere scenarija koji prikazuju okvir uključenosti v. okvir 2.5., a za detaljno objašnjenje tih pojmova Smjernice OECD-a za dubinsku analizu odgovornog poslovnog ponašanja.

SLIKA 2.1. OČEKIVANJA U POGLEDU DUBINSKE ANALIZE OVISNO O UKLJUČENOSTI U ŠTETNI UČINAK



Izvor: OECD (2018[3]), *Smjernice OECD-a za dubinsku analizu odgovornog poslovnog ponašanja*, https://rbcroatia.gov.hr/wp-content/uploads/2024/04/Smjernice-OECD-a-za-dubinsku-analizu-1.pdf.

OKVIR 2.5. SCENARIJI KOJI ILUSTRIRAJU OKVIR UKLJUČENOSTI

Izazivanje učinka

Poduzeće V razvilo je i uvelo generativni model umjetne inteligencije koji je sposoban stvarati tekstualne, audio i video izlazne rezultate na temelju upita korisnika. Model se trenira na osnovi podataka prikupljenih iz javno dostupnih izvora. Model se trenira i na osnovi osobnih podataka pojedinačnih korisnika koje poduzeće V prikuplja kad se korisnici pretplate na upotrebu modela i kad unose osobne podatke

u upite. Poduzeće V nije uspostavilo učinkovite zaštitne mehanizme kako bi korisnike informiralo o načinu na koji trenira model na osnovi njihovih podataka. Isto tako, ne onemogućuje curenje osobnih podataka u izlazne proizvode koje model generira.
Poduzeće V uzrokuje štetne učinke u pogledu prava na privatnost.

Doprinos štetnom učinku

Poduzeće X razvilo je nadzorni sustav utemeljen na umjetnoj inteligenciji i upravlja tim sustavom čija je namjena praćenje i analiza produktivnosti radnika, uključujući prikupljanje i analizu podataka putem kamera, elektroničke pošte, programa koji se koriste na prijenosnim računalima poduzeća te internih komunikacijskih alata za radnike. Nadzorni sustav utemeljen na umjetnoj inteligenciji ima mogućnost ukazivanja na određene vrste ponašanja radnika, kao što su česte negativne izjave o poduzeću, umor, agresija prema kolegama, zloupotreba podataka poduzeća itd. Poduzeće X prodaje taj sustav drugim poduzećima. Poduzeće A kupuje sustav umjetne inteligencije od poduzeća X i koristi ga za nadzor radnika u svojem distribucijskom skladištu. Otprije je poznato da je poduzeće A financijski kažnjavano radi nezakonitih aktivnosti suzbijanja sindikalnog organiziranja i praćenja radnika bez njihova informiranog pristanka. Informacije o raspoloženju zaposlenika koje prikuplja i analizira sustav umjetne inteligencije zatim se koriste pri donošenju odluka o raskidu ugovora zaposlenika za koje se smatra da se bave sindikalnom aktivnošću.
U tom scenariju, poduzeće X doprinosi štetnim učincima time što je u znatnoj mjeri olakšalo poduzeću A da izazove štetni učinak osmislivši sustav umjetne inteligencije čija bi namjena vjerojatno uzrokovala štetne učinke te nije uključilo dostatne zaštitne mjere u dizajn proizvoda. Nadalje je taj sustav ustupilo klijentu čija je dotadašnja praksa ukazivala na veću vjerojatnost nastanka štetnih učinaka.
U drugom, pak, scenariju koji uključuje isti sustav umjetne inteligencije, pristup sustavu umjetne inteligencije omogućen je jednoj akademskoj ustanovi kako bi ga koristila u istraživanju produktivnosti radnika. Međutim, i ta akademska ustanova sustav umjetne inteligencije koristi za nadzor ponašanja svojih zaposlenika bez njihova informiranog pristanka. U tom scenariju namjena sustava umjetne inteligencije koju je navela akademska ustanova nije upućivala na vjerojatnost da bi mogla prouzročiti štetan učinak niti se moglo razumno predvidjeti visoku rizičnost klijenta te stoga poduzeće X nije u znatnoj mjeri doprinijelo nastanku štetnih učinaka.

Izravna povezanost sa štetnim učinkom

Poduzeće Y radi na razvoju sustava umjetne inteligencije koji se koristi za otkrivanje određenog oblika raka kože. Sustav umjetne inteligencije trenira se na skupu podataka koji sadržava slike osoba oboljelih od raka kože. Skup podataka je kupljen od pružatelja usluga, poduzeća B. Poduzeće B razvija svoje skupove podataka tako da njegovi radnici pregledaju i označe tisuće slika, uključujući i eksplicitne slike. Mnogi od tih radnika prijavili su da su kao posljedicu tog posla osjetili štetne učinke na svoje mentalno zdravlje. Sjedište poduzeća B je u državi sa sustavom nadležnosti koji pruža slab stupanj zaštite radnika. Poduzeće B nema uspostavljene sustave koji bi podrazumijevali uključenost radnika niti pruža radnicima bilo kakve usluge mentalne podrške.
U ovom slučaju poduzeće Y nije uzročnik učinka na radnike poduzeća B niti takvom učinku doprinosi. Međutim, poduzeće Y je izravno povezano sa štetnim učincima koje uzrokuje poduzeće B jer je s poduzećem B u poslovnom odnosu.Kao što je prethodno navedeno, odnos poduzeća prema riziku nije statičan. Nastavi li poduzeće Y nabavljati usluge od poduzeća B ne poduzimajući korake za ublažavanje rizika, njegova izravna povezanost s rizikom s vremenom može prerasti u doprinos štetnom učinku.

Korak 2.4. – Odrediti prioritete najznačajnijih rizika

Na temelju pribavljenih informacija o stvarnim i potencijalnim štetnim učincima, pri poduzimanju daljnjih mjera odrediti redoslijed važnosti najznačajnijih rizika i štetnih učinaka prema njihovoj težini i vjerojatnosti da će nastupiti. Ne bude li moguće odmah pristupiti rješavanju svih potencijalnih i stvarnih štetnih učinaka, trebat će odrediti prioritete. Nakon što se prepoznaju i riješe najznačajniji štetni učinci, poduzeće bi trebalo pristupiti rješavanju manje značajnih predvidivih učinaka.
Gdje će rizik od štetnih učinaka biti najznačajniji, specifično je za poduzeće. Stoga je važno da poduzeća dokažu da provode vjerodostojan postupak određivanja prioriteta.
Uključivanje mjerodavnih dionika, uključujući radnike, predstavnike radnika i sindikate, u određivanje prioriteta te javno obrazlaganje načina na koji se donose odluke o određivanju prioriteta može doprinijeti izgradnji povjerenja u pristup poduzeća dubinskoj analizi. U nekim slučajevima određivanje prioriteta može proizlaziti iz nacionalnih pravnih obveza.

TABLICA 2.3. ČIMBENICI KOJE TREBA UZETI U OBZIR PRI UTVRĐIVANJU REDOSLIJEDA VAŽNOSTI RIZIKA

Razmjer	**Obuhvat**	**Nemogućnost otklanjanja (ili vraćanja u prvobitno stanje)**	**Vjerojatnost/ predvidivost**
Težina učinka	Doseg učinka	U kojoj se mjeri učinak može otkloniti (bilo kakvo ograničenje sposobnosti povratka pogođenih osoba ili okoliša na situaciju jednaku onoj u kojoj su bili prije štetnog učinka).	Procjena vjerojatnosti nastanka učinka
Primjeri učinaka znatnog razmjera Primjer: sustav umjetne inteligencije koristi se za određivanje trajanja kazni u kaznenim predmetima. Primjer: sustav umjetne inteligencije koristi se za generiranje eksplicitnih slika pojedinca u svrhu spolnog uznemiravanja i ucjene. Primjer: sustav umjetne inteligencije koristi se za prikupljanje informacija o ciljnoj skupini te praćenje njihovih kretanja i svakodnevnih navika radi olakšavanja djelovanja protiv njih.	Primjeri učinaka znatnog obuhvata Primjer: podatkovni centri koji se koriste za napajanje sustava umjetne inteligencije troše prekomjerne količine vode u zajednicama u kojima su lokalni vodni resursi ugroženi. Primjer: pristrane preporuke koje generira sustav umjetne inteligencije koji se koristi u državnim sustavima socijalne skrbi dovode do ukidanja financijske pomoći za tisuće obitelji.	Primjeri učinaka neotklonjive naravi Primjer: UI chatbot preporučuje pojedinom korisniku da počini samoozljeđivanje ili nanese štetu drugima.	Primjeri vjerojatnih ili predvidivih učinaka Primjer: brojna izvješća o tome kako zlonamjerni akteri zlorabe sustave umjetne inteligencije sličnih funkcija. Primjer: 20 % korisnika UI chatbota izvješćuje da chatbot tijekom razgovora dijeli nasilan sadržaj.

3. KORAK – ZAUSTAVLJATI, SPREČAVATI I UBLAŽAVATI ŠTETNE UČINKE

TABLICA 2.4. – 3. KORAK: PREGLED POVEZANIH ODREDBI U POSTOJEĆIM OKVIRIMA

ASEAN-OV VODIČ	Odjeljak C.2: Utvrđivanje razine uključenosti ljudi u odlučivanje potpomognuto umjetnom inteligencijom; C.3: Upravljanje operacijama; i Prilog A: 3.: Utvrđivanje razine uključenosti ljudi u odlučivanje potpomognuto umjetnom inteligencijom
Australske smjernice za uvođenje umjetne inteligencije (Prakse provedbe)	Praksa provedbe 1, 2, 3, 4, 5
Kanadski kodeks ponašanja	Sigurnosne mjere 2 i 3; Pravednost i jednakost, Transparentnost, Ljudski nadzor i praćenje, Valjanost i robusnost
HUDERIA	Plan ublažavanja učinaka i Pristup pravnim sredstvima
Akt EU-a o UI-ju	Uvodna izjava 115., čl. 9. st. 2.a, čl. 9. st. 2. t. (d), čl. 9. st. 4. – 5., čl. 50. (obveze transparentnosti), čl. 55. st. 1. t. (b)
DSA	Čl. 35.: Ublažavanje rizika
CSDDD	Članci 10. i 11.: Sprečavanje i (ako nije ili nije odmah moguće) ublažavanje potencijalnih negativnih učinaka te okončanje stvarnih negativnih učinaka i svođenje njihova razmjera na najmanju moguću razinu
Kodeks ponašanja Hirošimskog procesa	Načela 1., 2., 6. – 7. i 11.
IEEE 7000	10. Etički postupak projektiranja utemeljen na riziku
ISO 31000 i ISO/IEC 23894	6.5.: Postupanje s rizikom
ISO/IEC 42001	6.1.3. Postupanje s rizikom UI-ja; 8.3. Postupanje s rizikom UI-ja; Prilog A A.5. i kontrole A.5.2. – A.5.5. za upravljanje rizikom; A.7. (Podaci za sustave UI-ja) i kontrole A.7.2. – A.7.6. za kvalitetu podataka i upravljanje podacima

Japanske Smjernice za umjetnu inteligenciju u poslovanju	Dio 2.C. Zajednička vodeća načela, Dio 3., 4., 5.; Prilog 3., 4., 5.
Korejski Osnovni zakon o umjetnoj inteligenciji	Čl. 31., 32. i 34.
Singapurski okvir za testiranje AI Verify	Sigurnost 4.1.1. – 4.6.1. Zaštita 5.1.1. – 5.7.1. Robusnost 6.1.1. – 6.5.3.
Vodeća načela UN-a o poslovanju i ljudskim pravima	Operativno načelo 19.
Okvir za osiguravanje pouzdanosti UI-ja, UK DSIT	4.2. Mehanizmi osiguravanja UI-ja; 5.2. Spektar osiguranja UI-ja; 5.3. Osiguravanje podataka, modela, sustava i upravljanja u praksi; 6.1. Koraci za uspostavu osiguravanja UI-ja
Okvir SAD-a za upravljanje rizicima UI-a	Mapiranje 1, Upravljanje 1 – 4

KORAK 3.1. – PRISTUPANJE RJEŠAVANJU RIZIKA KOJE UZROKUJE ILI KOJIMA DOPRINOSI PODUZEĆE

Na temelju procjene poduzeća o njegovoj uključenosti u učinak prekinuti aktivnosti koje izazivaju štetne učinke ili im doprinose.
Izraditi i provesti planove za sprečavanje i ublažavanje potencijalnih (budućih) štetnih učinaka.

OKVIR 2.6. PRILAGODBA UPRAVLJANJA RIZICIMA OKOLNOSTIMA PODUZEĆA

Na prirodu i opseg dubinske analize mogu utjecati čimbenici poput konteksta poslovanja poduzeća te bi oni trebali biti razmjerni resursima poduzeća, njegovoj povezanosti sa štetnim učinkom i ozbiljnosti štetnog učinka. Velika poduzeća sa širokom paletom poslovnih djelatnosti i mnogim proizvodima ili uslugama možda će trebati formalnije i opsežnije sustave nego manja poduzeća s ograničenim rasponom proizvoda ili usluga kako bi učinkovito identificirala rizike i upravljala njima.

Postoje praktična ograničenja u pogledu načina na koji se od poduzeća očekuje da odgovore na rizike i učinke. Nastojanja da se rizici ublaže trebala bi biti razmjerna tim čimbenicima, vodeći računa o sljedećem:

- kontekst poslovanja poduzeća (npr. uloga poduzeća u lancu vrijednosti umjetne inteligencije, njegov utjecaj i tehničke mogućnosti)
- veličina poduzeća
- uključenost u štetni učinak (tj. da li poduzeće uzrokuje učinke, doprinosi im ili je s njima izravno povezano, v. okvir 2.5.)
- ozbiljnost štetnih učinaka.

Svako poduzeće treba dati svoj doprinos, a odgovornost jednog poduzeća ne bi se trebala prebacivati na druga. Isto tako, dubinska analiza odgovornog poslovnog ponašanja nije standard savršenosti, nego standard poboljšanja. Od poduzeća se ne očekuje da odmah riješe sve rizike i učinke u koje su uključena. Umjesto toga, poduzeća bi trebala nastojati postupno poboljšavati svoje sustave i procese kako bi izbjegavala štetne učinke i poduzimala mjere za njihovo otklanjanje.

Praktični primjeri provedbe (za sva poduzeća u lancu vrijednosti umjetne inteligencije, skupine 1 – 3)

1. Dodijeliti odgovornost mjerodavnom rukovodećem osoblju kako bi se osigurao prestanak aktivnosti koje izazivaju štetne učinke ili im doprinose te kako bi se spriječile aktivnosti koje bi mogle imati štetne učinke u budućnosti ili im doprinijeti. Ovisno o kontekstu rizika, to može uključivati osoblje iz različitih poslovnih jedinica koje raspolaže sredstvima za poduzimanje potrebnih mjera za pristupanje rješavanju rizika (npr. istraživanje i razvoj proizvoda, nabava, upravljanje odnosima s klijentima, prodaja ili pravni poslovi).
2. U slučaju složenih aktivnosti koje je teško prekinuti zbog operativnih, ugovornih ili pravnih razloga (npr. pružanje javnih usluga, dugoročni ugovori, ovisnost o strani u poslovnom odnosu), izraditi plan postupnog prestanka aktivnosti koje uzrokuju štetne učinke ili im doprinose. Poduzećima može ići u prilog ako javnosti objasne složenost situacije i trud koji ulažu u to da s vremenom postupno zaustave aktivnosti.
3. Savjetovati se i uključiti dionike izložene štetnim učincima i one koji bi im mogli postati izloženi, kao i njihove predstavnike, radi osmišljavanja odgovarajućih mjera i provedbe plana (v. poglavlje 1., Svrhovita uključenost dionika).

4. Na temelju rezultata procjene rizika ažurirati i jačati sustave upravljanja kako bi se osiguralo bolje praćenje podataka i ukazalo na rizike prije no što nastupe štetni učinci.
5. Ažurirati politike poduzeća, uz aktivno uključivanje dionika, kako bi se pružile smjernice za izbjegavanje i borbu protiv štetnih učinaka u budućnosti te osigurala njihova provedba.
6. Pri odlučivanju o mjerama ublažavanja, prema potrebi usporediti te mjere s drugim mogućim rizicima koji bi mogli nastupiti, kao i mogućim koristima od uvođenja ili upotrebe sustava umjetne inteligencije.

OKVIR 2.7. KORIŠTENJE UMJETNE INTELIGENCIJE ZA PODRŠKU DUBINSKOJ ANALIZI ODGOVORNOG POSLOVNOG PONAŠANJA

Osim u okviru lanaca vrijednosti umjetne inteligencije, umjetna inteligencija može osigurati podršku i pri dubinskoj analizi odgovornog poslovnog ponašanja. Primjerice,

- pomagati u analizi velikih količina podataka o dobavljačima radi prepoznavanja potencijalnih štetnih učinaka u složenim globalnim lancima vrijednosti koje bi bilo teško ili skupo pratiti ručno
- pregledavati vijesti, izvješća i društvene mreže kako bi pravodobno upozorila na nove probleme povezane s poslovanjem poduzeća ili stranama s kojima je ono u poslovnim odnosima te pružila ciljane informacije o usklađenosti više sustava nadležnosti, pomažući multinacionalnim poduzećima da se lakše snalaze u promjenjivim okvirima na različitim tržištima
- pomagati pri provjeri podrijetla proizvoda i uvjeta proizvodnje, čime se povećava transparentnost i omogućava učinkovitije otklanjanje prepoznatih problema.

Praktični primjeri provedbe (za poduzeća uključena u životni ciklus sustava umjetne inteligencije, skupina 2)

Mjere kojima se izbjegava i ublažava rizik od toga da sustav umjetne inteligencije izazove štetne učinke ili im doprinese mogu se u širem smislu razvrstati u četiri skupine navedene u nastavku, no odgovarajuće će mjere ovisiti o konkretnom sustavu umjetne inteligencije i slučaju upotrebe te se može razmotriti niz dodatnih mjera:

- **odgovorno pribavljanje i upotreba podataka za treniranje modela umjetne inteligencije**, uključujući tijekom faze prikupljanja i obrade podataka životnog ciklusa sustava umjetne inteligencije (npr. mjere za procjenu i poboljšanje kvalitete podataka i učinkovitosti sustava umjetne inteligencije te sprečavanje ili ublažavanje rizika pribavljanja podataka prikupljenih ili označenih na načine koji uzrokuju štetne učinke)
- **transparentnost, objašnjivost i sljedivost,** osobito nakon uvođenja (npr. mjere za informiranje dionika o funkcijama, sposobnostima i rizicima sustava umjetne inteligencije)
- **sigurnost, uključujući fizičku i kibernetičku sigurnost te robusnost** tijekom životnog ciklusa sustava umjetne inteligencije (npr. mjere za osiguravanje otpornosti sustava umjetne inteligencije na napade te njegove pouzdanosti, ponovljivosti, reproduktivnosti i predvidljivosti)
- **odgovorno uvođenje**, uključujući odgovorno upravljanje i praćenje te, ako je to primjereno, povlačenje iz upotrebe (npr. mjere kojima se procjenjuje je li model siguran za uvođenje te primjena odgovarajućih zaštitnih mehanizama).

Odgovorno pribavljanje podataka i treniranje

1. Provoditi preglede kvalitete podataka radi prepoznavanja i rješavanja problema poput netočnih oznaka i reprezentativnosti.
2. Uvesti pristupe prikupljanju podataka i treniranju sustava umjetne inteligencije koji čuvaju privatnost i osiguravaju odgovorno upravljanje podacima, kao što su čišćenje podataka, obrada na uređaju i federalno učenje. Pratiti prethodno trenirane modele koji se koriste za razvoj u okviru redovitog praćenja i održavanja sustava umjetne inteligencije, uključujući putem pregleda kvalitete podataka.
3. Ako poduzeće nije sigurno u to hoće li moći utrenirati siguran model u prvotno planiranom opsegu, može razmotriti postupno povećavanje opsega (tj. treniranje manjeg ili u drugom smislu slabijeg modela).

4. Primjenjivati najsuvremenije tehnike usklađivanja i sigurnosti, kao što je obrnuto podržano učenje (Centar za upravljanje UI-jem, 2023[19]).
5. Poduzeti korake za sprečavanje ili ublažavanje rizika povezanih s prikupljanjem i obradom podataka.
6. Rizici u vezi s kvalitetom i pribavljanjem podataka mogu biti povezani i s uslugama obogaćivanja podataka.

Transparentnost, objašnjivost i sljedivost

7. Nastojati omogućiti transparentnost, objašnjivost i sljedivost u vezi s pribavljanjem podataka od podugovaratelja, skupovima podataka, postupcima, mjerodavnim odlukama donesenima tijekom razvoja sustava, uključujući preispitivanje značajnih odluka koje provodi čovjek, kao i žalbene postupke (v. okvir 2.8.).

OKVIR 2.8. OMOGUĆAVANJE TRANSPARENTNOSTI, OBJAŠNJIVOSTI I SLJEDIVOSTI TIJEKOM ŽIVOTNOG CIKLUSA SUSTAVA UMJETNE INTELIGENCIJE

Transparentnost se u ovom kontekstu odnosi na objavljivanje informacija kako bi ljudi bili svjesni da se umjetna inteligencija koristi u predviđanju, preporuci, odlučivanju ili interakciji (npr. chatbot). Transparentnost isto tako znači omogućiti ljudima da razumiju kako se sustav umjetne inteligencije razvija, trenira, kako radi i kako se uvodi u domeni primjene, kako bi, primjerice, korisnici i potrošači mogli donositi informiranije odluke. Ovisno o kontekstu i osim ako to ne zahtijeva zakon, transparentnost se ne mora proširiti na objavljivanje izvornog kôda ili drugih vlasničkih kôdova ili skupova podataka, koji bi mogli biti previše tehnički složeni da bi bili korisni za razumijevanje ishoda.

Objašnjivost znači omogućiti dionicima da razumiju kako se određuje ishod sustava umjetne inteligencije. To podrazumijeva pružanje lako razumljivih informacija koje onima koji su pogođeni štetnim učincima

mogu omogućiti da ospore ishod, osobito – u mjeri u kojoj je to izvedivo – čimbenike i logiku koji su doveli do ishoda.

Objašnjivost se može ostvariti na različite načine, ovisno o kontekstu (kao što je značaj ishodâ). Primjerice, za neke vrste sustava umjetne inteligencije, zahtijevanje objašnjivosti može negativno utjecati na točnost i učinkovitost sustava (jer može zahtijevati smanjivanje varijabli rješenja na skup dovoljno malen da ga ljudi mogu razumjeti, što možda neće biti optimalno u složenim, visokodimenzionalnim problemima), ili na privatnost i sigurnost. Može i povećati složenost i troškove, potencijalno dovodeći aktere umjetne inteligencije koji su MSP-ovi u nerazmjerno nepovoljan položaj.

Stoga, kad akteri umjetne inteligencije objašnjavaju ishod, mogu razmotriti da – jasnim i jednostavnim jezikom te primjereno kontekstu – navedu glavne čimbenike u odlučivanju, odlučujuće čimbenike, podatke, logiku ili algoritam na kojem se temelji konkretan ishod ili objasne zašto su naizgled slične okolnosti dovele do različitog ishoda. To treba učiniti tako da se pojedincima omogući da razumiju i ospore ishod, prema potrebi poštujući obveze zaštite osobnih podataka

Iz perspektive razvoja, objašnjivost se smatra ključnom za učenje iz pogrešaka sustava. Bez nje se ne mogu steći vrijedni uvidi na osnovi pogrešaka. Sposobnost razumijevanja postupaka odlučivanja koji su doveli do pogrešaka smatra se ključnom za poboljšanje sustava umjetne inteligencije i izgradnju povjerenja.

Sljedivost u području umjetne inteligencije opisuje nastojanje da se vodi evidencija o podrijetlu podataka, procesa, kôda i drugih elemenata u razvoju sustava umjetne inteligencije. Sljedivost često obuhvaća detaljne informacije o elementu ili komponenti sustava umjetne inteligencije, kao što su ulazni podaci ili model. Ključna je za omogućavanje revizije sustava. Radi omogućavanja transparentnosti i sljedivosti, poduzeća mogu dokumentirati informacije nabrojane u nastavku ako su one relevantne te izvedive u odnosu na njihovu konkretnu aktivnost tijekom životnog ciklusa sustava umjetne inteligencije:

- informacije o upotrebi i rizicima prepoznate u koraku 2.1.
- izvori podataka, postupci prikupljanja podataka i informacije o obradi podataka
- cjelokupan kôd, uključujući potrebne biblioteke
- informacije o načinu izvršavanja kôda radi osiguravanja reproduktivnosti izlaznih podataka, uključujući detaljnu dokumentaciju o parametrima i računalnim zahtjevima
- informacije o tome na koji način se upotrebljavaju izlazni podaci

modela te generiraju li sustavi umjetne inteligencije izlazne podatke ili odluke (npr. objavljivanje informacija o slikama, zvuku ili tekstu koje generira umjetna inteligencija)
- informacije o strategiji praćenja, uključujući pokazatelje uspješnosti, pragove, očekivano ponašanje modela i mjere ublažavanja; informacije o nedostacima, ograničenjima i pristranostima modela, kao i o tome prenose li se te informacije mjerodavnim dionicima i ako da, na koji način.

Objavljivanje informacija treba prilagoditi njihovim primateljima, što može zahtijevati više načina objavljivanja s različitim stupnjevima detaljnosti (npr. oznake umjetne inteligencije o nutritivnoj vrijednosti, podatkovni listovi, kartice modela, kartice sustava, tehnička izvješća itd.).

Izvori: OECD (2024[20]), AI, data governance and privacy: Synergies and areas of international co-operation, https://doi.org/10.1787/2476b1a4-en; Ministarstvo trgovine SAD-a. Nacionalna uprava za telekomunikacije i informacije (2024[21]), Izvješće o politici odgovornosti za umjetnu inteligenciju, https://www.ntia.gov/sites/default/files/publications/ntia-ai-report-final.pdf; Europska unija (2024[14]), Uredba (EU) 2024/1689 Europskog parlamenta i Vijeća od 13. lipnja 2024. o utvrđivanju usklađenih pravila o umjetnoj inteligenciji, https://eur-lex.europa.eu/legal-content/HR/TXT/HTML/?uri=OJ:L_202401689.

8. Nastojati omogućiti načine generiranja i osiguravanja interpretacija i objašnjenja izlaznih podataka sustava umjetne inteligencije, prema potrebi uključujući u objašnjenja modela sljedeće informacije:
   a) vrsta i izvor ulaznih podataka modela
   b) postupak transformacije glavnih podataka
   c) kriteriji odlučivanja i njihovo obrazloženje
   d) objava o korištenju umjetne inteligencije.
9. Poduzeća trebaju uvesti mehanizme za pružanje jasnih, dostupnih i smislenih objašnjenja automatiziranih postupaka odlučivanja, osobito kad takve odluke mogu znatno utjecati na pojedince. Ta objašnjenja trebaju uključivati logiku, glavne parametre i moguće ishode algoritamskog postupka, prilagođene znanju prosječnog korisnika.
10. Razviti i uvesti pouzdane mehanizme autentifikacije sadržaja i utvrđivanja njegova podrijetla, gdje je to tehnički izvedivo (v. okvir 2.9.).

OKVIR 2.9. MEHANIZMI AUTENTIFIKACIJE SADRŽAJA I UTVRĐIVANJA PODRIJETLA

Podrijetlo se općenito odnosi na osnovne, pouzdane činjenice o nastanku određenog digitalnog sadržaja (slika, videozapis, audiozapis, dokument). Može uključivati informacije o tome tko ga je izradio te kako, kada i gdje je izrađen ili uređen. Utvrđivanje podrijetla digitalnog sadržaja trenutačno je iznimno teško s obzirom na razmjer i brzinu interneta jer je softver za manipulaciju sve sofisticiraniji, a metapodaci se isto tako mogu lako manipulirati te ne pružaju dokaz o podrijetlu. Trenutno su u izradi tehnička rješenja i dobre prakse u okviru zajedničkih nastojanja struke i dionika, kao što su Smjernice za provoditelje *(Guidance for Implementers)* i Smjernice za umjetnu inteligenciju i strojno učenje *(Guidance for AI and ML)* Koalicije za podrijetlo i autentičnost sadržaja (C2PA), kao i Odgovorne prakse za sintetičke medije *(Responsible Practices for Synthetic Media)* Partnerstva za UI (PAI). Općenito govoreći, postojeća dobra praksa u ovom području uključuje sljedeća tri glavna elementa:

- transparentnost u pogledu sposobnosti, funkcija, ograničenja i mogućih rizika tehnologija koje proizvode sintetičke medije
- integracija izravnih ili neizravnih metoda objavljivanja (npr. oznake sadržaja, vodeni žigovi)
- ulaganje u istraživanje i razvoj metoda otkrivanja i trajnosti kriptografskog označavanja.

Izvori: OECD (2022[13]), *OECD Framework for the Classification of AI systems*, https://doi.org/10.1787/cb6d9eca-en; Coalition for Content Provenance and Authenticity (C2PA) (n.d.[22]), Guiding Principles for C2PA Designs and Specifications, https://c2pa.org/principles/; Partnership on AI (n.d.[23]), Responsible Practices for Synthetic Media, https://syntheticmedia.partnershiponai.org/#read_the_framework.

11. Razmotriti sudjelovanje u unapređivanju i normizaciji znanosti o mjerenju umjetne inteligencije radi potpunog razumijevanja dugoročnih koristi i rizika sustava umjetne inteligencije.
12. Razmotriti izradu vodiča za vanjske dionike kao resursa koji bi dionicima omogućio da bolje razumiju sustav umjetne inteligencije.

Taj bi vodič mogao biti u obliku dokumentacije o odgovornom UI-ju kao jedinstvene baze podataka u kojoj dionici mogu pronaći informacije o namjenama i ograničenjima, odabirima u projektiranju odgovorne umjetne inteligencije te dobrim praksama njezina uvođenja i optimizacije njezine učinkovitosti. Mogao bi obuhvatiti pitanja povezana s OECD-ovim Načelima za umjetnu inteligenciju. Dokument bi se kontinuirano razvijao sa stjecanjem dubljih spoznaja o sustavu umjetne inteligencije i povezanim rizicima.

13. Informacije u objavama učiniti dovoljno jasnima i razumljivima kako bi se subjektima koji uvode sustav i korisnicima, prema potrebi te ovisno o mjerodavnosti, omogućilo tumačenje izlaznih podataka modela/sustava te kako bi se korisnicima omogućila odgovarajuća upotreba, pri čemu objave trebaju biti potkrijepljene i utemeljene na robusnim postupcima dokumentiranja.

Sigurnost, zaštita i robusnost

1. Osmisliti pristupe kojima će se osigurati podrška robusnosti, sigurnosti i zaštiti tijekom cjelokupnog životnog ciklusa sustava umjetne inteligencije, primjerice na sljedeće načine:
   a) provođenje penetracijskog testiranja u različitim fazama životnog ciklusa sustava umjetne inteligencije, razmjerno riziku
   b) praćenje ponašanja sustava umjetne inteligencije, uključujući mehanizme za prikupljanje i procjenu ulaznih podataka korisnika i drugih mjerodavnih organizacija u području umjetne inteligencije, žalbe i nadjačavanja, povlačenje iz upotrebe, odgovor na incidente, oporavak i upravljanje promjenama
   c) uspostava mehanizama brzog odgovora na kvarove sustava umjetne inteligencije, kao što su mehanizmi za zamjenu, isključivanje ili deaktivaciju sustavâ čija učinkovitost ili ishodi nisu u skladu s njihovom namjenom
   d) uspostava robusnog programa za otkrivanje prijetnji iznutra
   e) osiguravanje težinskih faktora modela
   f) praćenje promjena učinkovitosti u odnosu na dogovorene pokazatelje tijekom životnog ciklusa sustava umjetne inteligencije
   g) uspostava kibernetičkosigurnosne zaštite
   h) praćenje ishoda sustava umjetne inteligencije radi utvrđivanja pomaka u podacima i modelu.

Odgovorno uvođenje

1. Uključiti dionike radi prikupljanja ulaznih elemenata o zahtjevima sustava i odlukama o dizajnu (npr. „sustav treba poštovati privatnost svojih korisnika”) prije uvođenja sustava (v. okvir 2.10.).

OKVIR 2.10. PLAN ODGOVORA PRIJE UVOĐENJA SUSTAVA

Jasan i transparentan plan odgovora ključan je element za sprečavanje rizika pri uvođenju sustava umjetne inteligencije. U tom se planu mogu jasno definirati sljedeći aspekti:

- scenariji rizika koji nalažu korekcije pri uvođenju sustava, razrađeni u postupku modeliranja prijetnji, te signali koji ukazuju na odstupanja od očekivanog ponašanja
- tim za odgovor koji bi u svom sastavu imao predstavnike IT-a, kibernetičke sigurnosti, razvoja umjetne inteligencije, pravnih poslova, komunikacija i mjerodavnih poslovnih jedinica, kao i vanjske stručnjake za određena područja. Kako su mogući raznoliki scenariji rizika, odgovor na incidente može iziskivati stručno znanje koje nadilazi stručnost samih razvojnih inženjera umjetne inteligencije te *inpute* više strana
- uloge i odgovornosti različitih timova i pojedinaca uključenih u postupak odgovora na incidente. Radi brzog djelovanja, svi članovi tima trebaju biti upoznati s odgovornostima i odlukama u svojoj nadležnosti
- razina do koje je odlučivanje automatizirano odnosno prepušteno ljudskim operaterima
- razina do koje su ovlasti podijeljene
- razina do koje su protokoli za korekciju pri uvođenju sustava obvezujući.

Za više pojedinosti v. Institut za politiku i strategiju UI-ja 2023[24], ISO (2023[25]).

2. Razmotriti mjere za postupno uvođenje sustava umjetne inteligencije kako se pojavljuju dokazi o rizicima. Istraživanja su prepoznala „gradijentni sustav pristupa” uvođenju generativnih modela umjetne inteligencije, koji se kreće od potpuno zatvorenog i

postupnog/faznog objavljivanja na jednom kraju gradijenta do preuzimanja i potpune otvorenosti na drugom kraju (Solaiman, 2023[26]). Svaka razina pristupa nosi rizike i kompromise koje bi trebalo uzeti u obzir pri razmatranju načina sprečavanja i ublažavanja rizika povezanih sa sustavom umjetne inteligencije (v. okvir 2.11.).

OKVIR 2.11. SPREČAVANJE ILI UBLAŽAVANJE RIZIKA PRI UVOĐENJU SUSTAVA UMJETNE INTELIGENCIJE

U istraživanjima su prepoznati konkretni tehnički alati u koje poduzeća mogu ulagati, kao i netehničke mjere koje mogu poduzeti kako bi pristupila pitanju rizika pri uvođenju generativnih sustava umjetne inteligencije. Njihova se relevantnost razlikuje ovisno o sustavu i scenariju uvođenja te se stoga navode kao primjeri mogućih alata i mjera koje poduzeća mogu razmotriti.

Neki od tehničkih alata su:

- ograničavanje učestalosti – ograničavanje broja izlaznih rezultata koje korisnik može generirati
- sigurnosni filtri i filtri sadržaja – filtri razvijeni tako da pokrenu prazan odgovor kad prime potencijalno nesiguran ulaz
- modeli za detekciju – tehnička i ljudska detekcija sadržaja koji generira umjetna inteligencija
- unaprijed zadani odgovori – unaprijed određeni sigurni izlazni rezultati koji se aktiviraju na određene ulazne elemente, a mogu se ugraditi u sučelje modela.

Neke od netehničkih mjera su:

- interne politike upravljanja rizicima i kodeksi ponašanja (v. 1. korak)
- uspostava pravnih zaštitnih mehanizama putem licenci, pri čemu nositelj licence može provoditi sankcije za kršenje uvjeta upotrebe. To je moguće kod restriktivnijih tipova sustava umjetne inteligencije, ali potencijalno teško kod sustava koji su u potpunosti dostupni za preuzimanje ili otvoreni
- ulaganja u osposobljavanje o ljudskim pravima i radnim pravima za osobe koje nisu stručnjaci, a sudjeluju u projektiranju i uvođenju sustava umjetne inteligencije
- ulaganja u aktivnosti predviđanja rizika.

Izvor: OECD (2023[17]), Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI, https://doi.org/10.1787/2448f04b-en.

3. Razmotriti praćenje i kontrolu upotrebe modela ili sustava (npr. prikupljanjem informacija u okviru postupka upoznavanja klijenta (KYC) i ograničavanjem pristupa sustavu ili nekim njegovim funkcijama). To je moguće u okviru pristupa putem ograničavanja i eskalacije utemeljenih na riziku (v. (BSR, 2022[27])).
4. Osmisliti odgovarajuće interne procjene i mjere praćenja te pružiti potporu vanjskim znanstvenicima koji raspolažu resursima i znanjem za provedbu procjena nakon uvođenja sustava.
5. Uspostaviti i integrirati postupke prikupljanja povratnih informacija za krajnje korisnike i mjerodavne dionike radi prijave problema i ulaganja žalbi na ishode sustava.
6. Nakon uvođenja, ažurirati i prilagođavati sustav umjetne inteligencije na temelju njegova kontinuiranog praćenja i procjenjivanja. Metode prilagodbe uključuju podržano učenje uz povratne informacije čovjeka ili prilagođavanje skupova podataka.
7. Ako postoji prijetnja od značajnih štetnih učinaka ili su već počele nastajati ozbiljne štete, važno je na odgovoran način prekinuti razvoj i uvođenje sustava umjetne inteligencije dok se ne uspostavi dovoljna kontrola nad rizicima.
8. Procijeniti mogu li učinke prouzročiti pravni zahtjevi u pogledu razvoja i upotrebe umjetne inteligencije (v. okvir 2.12.).

OKVIR 2.12. UVOĐENJE U KONTEKSTIMA U KOJIMA SU ZAKONI NEUSKLAĐENI S MEĐUNARODNIM STANDARDIMA O ODGOVORNOM POSLOVNOM PONAŠANJU

U kontekstima u kojima nacionalni pravni zahtjevi mogu biti u suprotnosti s međunarodnim standardima odgovornog poslovnog ponašanja, poduzeća bi trebala naširoko jasno isticati svoju predanost poštivanju međunarodno priznatih ljudskih prava.
Tu predanost treba, kao preventivnu mjeru, jasno istaknuti i dogovoriti prije uvođenja sustava. Ako se pravni kontekst promijeni, poticati vlade da ispunjavaju svoje obveze iz područja ljudskih prava, osobito kad postoje izravne poveznice s poslovanjem poduzeća. Izbjegavati doprinos nepravednoj kriminalizaciji branitelja ljudskih prava ili upotrebi sustava umjetne inteligencije za suzbijanje mirnih prosvjeda. Razmotriti odustajanje od ulaska ili povlačenje iz konteksta u kojima se ljudska prava ne mogu poštovati.

9. Primjeri mjera koje poduzeća mogu koristiti za ublažavanje rizika zloupotrebe sustava umjetne inteligencije uključuju sljedeće:
   a) ograničenja na razini korisnika (npr. razmotriti odbijanje suradnje s određenim korisnicima ili skupinama kad je rizik zloupotrebe značajan)
   b) ograničenja učestalosti pristupa (npr. ograničavanje broja izlaznih podataka koje sustav može generirati po satu)
   c) ograničenja sposobnosti ili funkcija (npr. filtriranje izlaznih podataka ili smanjenje kontekstualnog prozora sustava)
   d) ograničenja slučajeva upotrebe (npr. zabrane određenih primjena sustava u određenim kontekstima)
   e) privremena ili trajna obustava rada sustava umjetne inteligencije (v. okvir 2.13.).

OKVIR 2.13. PRIVREMENA ILI TRAJNA OBUSTAVA RADA SUSTAVA UMJETNE INTELIGENCIJE

Protokoli obustave mogu uključivati postupak odobravanja ponovnog uvođenja sustava ili alternativne planove oporavka. Važno je da taj postupak oporavka prođe opsežno testiranje i validaciju, po mogućnosti uz uključivanje vanjskih dionika. Za ponovno uvođenje modela

za koji je dokazano da može izazvati štetne učinke trebalo bi postaviti visok prag. Ako ispravci nisu mogući ili nisu dovoljno robusni, umjesto ponovnog uvođenja sustava trebalo bi provoditi alternativne planove (npr. stavljanje modela izvan upotrebe i/ili koordinacija s drugim akterima iz državnih ili strukovnih tijela radi upravljanja odgovorima na razini struke). U ekstremnim slučajevima oporavak možda neće biti moguć.

Izvor: OECD (2025[28]), Towards a common reporting framework for AI incidents, https://doi.org/10.1787/f326d4ac-en.

10. Poduzeti korake kako bi se osiguralo da strane u poslovnim odnosima koje uvode sustave umjetne inteligencije poduzeća uoči samog uvođenja također svrhovito uključe dionike, osobito na radnom mjestu. V. korak 3.2. – Pristupanje rješavanju rizika izravno povezanih s poduzećem u cijelom lancu vrijednosti umjetne inteligencije.

Praktični primjeri provedbe (za korisnike sustava umjetne inteligencije, skupina 3)

1. U kontekstu upotrebe sustava umjetne inteligencije za donošenje operativnih odluka poduzeća će radi ublažavanja rizika možda morati uključiti radnike, predstavnike radnika i sindikate. Uključivanje radnika može obuhvaćati:
   a) objavljivanje informacija o sustavu umjetne inteligencije, uključujući njegov dizajn i namjenu
   b) razvoj mehanizama upravljanja kojima bi se radnicima omogućio pristup prikupljanju i analizi podataka koji se na njih odnose te ostvarivanje prava u tom pogledu
   c) osmišljavanje ili održavanje mehanizama za podnošenje pritužbi za radnike na koje utječe umjetna inteligencija
   d) osmišljavanje programa prekvalifikacije i programa pismenosti u području umjetne inteligencije.
2. Pri upotrebi sustava umjetne inteligencije u proizvodima i uslugama poduzeća mogu, gdje je to tehnički izvedivo, provoditi neovisna testiranja (npr. penetracijska ili druga) radi provjere kvalitete izlaznih podataka i ispitivanja ranjivosti. Poduzeća bi trebala razmotriti i objavljivanje informacija o tome generiraju li se izlazni podaci uz pomoć sustava umjetne inteligencije i u kojoj mjeri (v. okvir 2.8.).

Korak 3.2. – Pristupanje rješavanju rizika koji su izravno povezani s poduzećem duž cijelog lanca vrijednosti umjetne inteligencije

Na temelju određivanje prioriteta rizika izraditi i provoditi planove sprečavanja ili ublažavanja stvarnih ili potencijalnih štetnih učinaka izravno povezanih s poduzećem putem strana u poslovnim odnosima (npr. privremena obustava odnosa, nastavak odnosa tijekom provedbe mjera ublažavanja rizika ili raskid poslovnih odnosa).

Tijekom razvoja i upotrebe umjetne inteligencije, poduzeća mogu biti izravno povezana sa štetnim učincima koje uzrokuju 1.) drugi akteri umjetne inteligencije u životnom ciklusu sustava umjetne inteligencije ili 2.) strane u poslovnim odnosima izvan njihova životnog ciklusa, kao što su dobavljači ulaznih elemenata za umjetnu inteligenciju i korisnici sustava umjetne inteligencije.

Odgovarajući odgovori na rizike povezane sa stranama u poslovnim odnosima mogu ponekad uključivati sljedeće:

- **nastavak odnosa** tijekom provedbe mjera ublažavanja rizika
- **privremenu obustavu odnosa** uz istodobno provođenje mjera ublažavanja rizika
- **raskid poslovnih odnosa** nakon neuspješnih pokušaja ublažavanja ili kad poduzeće procijeni da ublažavanje nije izvedivo ili zbog ozbiljnosti štetnog učinka. Pri odluci i naknadnom planu o raskidu poslovnih odnosa trebalo bi uzeti u obzir moguće štetne učinke na društvo, okoliš i gospodarstvo, kao i svrhovito uključivanje dionika. U tim bi planovima trebalo detaljno opisati mjere koje će organizacija poduzimati, kao i njezina očekivanja od dobavljača, kupaca i drugih strana u poslovnim odnosima (v. okvir 2.15.). Mjere raskida poslovnih odnosa trebale bi biti u skladu s primjenjivim zakonima, uključujući pravo tržišnog natjecanja.

Praktični primjeri provedbe (za sva poduzeća u lancu vrijednosti umjetne inteligencije, skupine 1 – 3)

1. Dodijeliti odgovornost za razvoj, provedbu i praćenje planova sprečavanja ili ublažavanja stvarnih ili potencijalnih štetnih učinaka koji su izravno povezani s poduzećem putem strana u poslovnim odnosima.
2. Podupirati mjerodavne strane u poslovnim odnosima ili surađivati s njima u izradi prilagođenih planova kako bi se spriječili ili ublažili štetni učinci utvrđeni u razumnim i jasno definiranim rokovima, koristeći kvalitativne i kvantitativne pokazatelje za definiranje i mjerenje poboljšanja (takozvani „korektivni akcijski planovi").

3. U mjeri u kojoj je to moguće te u skladu s obvezama iz prava tržišnog natjecanja nastojati utjecati na strane u poslovnim odnosima kako bi ih se potaknulo na sprečavanje ili ublažavanje štetnih učinaka ili rizika. Nakon što se proizvod ili usluga prodaju ili preprodaju, razmotriti načine na koje bi bilo moguće ostvarivati utjecaj ograničavanjem pružanja ključnih usluga o kojima ovisi rad sustava umjetne inteligencije (npr. korisnička podrška, ažuriranja, poslužitelji itd.). Korištenje utjecaja može uključivati sljedeće:
   a) suradnju sa stranom u poslovnom odnosu kako bi je se potaknulo da spriječi i/ili ublaži učinak putem izravne komunikacije s osobljem odgovornim za pristupanje rješavanju rizika na operativnoj razini, razini višeg rukovodstva i/ili uprave, radi izražavanja mišljenja o tematskim područjima odgovornog poslovnog ponašanja
   b) uvrštavanje u komercijalne ugovore očekivanja koja se odnose na odgovorno poslovno ponašanje i dubinsku analizu
   c) povezivanje poslovnih poticaja – kao što je obvezivanje na dugoročne ugovore i buduće narudžbe – s uspjehom u području odgovornog poslovnog ponašanja
   d) suradnju s regulatornim tijelima i donositeljima politika o tematskim područjima odgovornog poslovnog ponašanja kako bi oni mogli utjecati na promjenu protuzakonitih praksi subjekta koji uzrokuje štetu
   e) priopćavanje – javno ili privatno – mogućnosti raskida poslovnih odnosa u slučaju neispunjavanja očekivanja u pogledu odgovornog poslovnog ponašanja.
4. Ako utjecaj poduzeća nije dovoljan da bi potaknuo strane u poslovnim odnosima da spriječe ili ublaže štetni učinak, u okvirima prava tržišnog natjecanja razmotriti načine za uspostavljanje dodatnog utjecaja na strane u poslovnim odnosima, prema potrebi uključujući, primjerice, više rukovodstvo, kao i podršku i poticaje.
5. U mjeri u kojoj je to moguće, u okvirima prava tržišnog natjecanja surađivati s drugim poduzećima ili dionicima radi izgradnje i ostvarivanja utjecaja kako bi se potaknulo sprečavanje i ublažavanje štetnih učinaka, primjerice surađujući unutar strukovnih udruženja ili s vladama.
6. Kako bi se spriječili potencijalni (budući) štetni učinci i riješili stvarni učinci, nastojati potaknuti nove i postojeće poslovne odnose, primjerice kroz politike ili kodekse ponašanja, ugovore ili pisane sporazume (v. okvir 2.14.).

OKVIR 2.14. POSEBNA RAZMATRANJA ZA PODUZEĆA KOJA SURAĐUJU S „KONTROLNIM TOČKAMA"

U kontekstu dubinske analize odgovornog poslovnog ponašanja, kontrolne točke predstavljaju ključne točke vidljivosti, utjecaja ili transformacije u razvoju proizvoda ili trgovini njime. Pod pretpostavkom da je to u skladu s pristupom utemeljenim na riziku te s nacionalnim pravom, uključujući pravo tržišnog natjecanja, usmjeravanje dubinske analize na kontrolne točke doprinosi učinkovitosti ukupnih aktivnosti dubinske analize.

Obilježja kontrolnih točaka općenito su, među ostalim, 1.) relativno mali broj aktera u lancu vrijednosti koji su u izravnim ili neizravnim poslovnim odnosima s mnogim drugim poduzećima i 2.) činjenica da podliježu (ili će uskoro podlijegati) regulativi i reviziji. U kontekstu razvoja i upotrebe sustava umjetne inteligencije, neka velika poduzeća mogu ispunjavati mnoga obilježja „kontrolnih točaka" (npr. proizvođači poluvodiča, temeljni modeli i vrlo velike internetske platforme). Dubinska analiza kontrolnih točaka kako bi se utvrdilo provode li i same dubinsku analizu pruža određenu razinu sigurnosti u pogledu toga da su rizici prepoznati, spriječeni i ublaženi, a da nema potrebe za detaljnom dubinskom analizom svih drugih odgovarajućih poduzeća. Osim toga, kontrolne točke obično već podliježu reviziji, obvezi javnog izvješćivanja ili nekom vidu regulatornog nadzora, što upućuje na to da usmjeravanje dubinske analize na tu točku može doprinijeti izbjegavanju nepotrebnog ponavljanja aktivnosti.*

Prepoznavanje i uključivanje kontrolnih točaka može se provoditi uključivanjem, u ugovore s dobavljačima i drugim stranama u poslovnim odnosima, zahtjeva da se prepoznaju kontrolne točke i da one ispune očekivanja u pogledu dubinske analize, upotrebom povjerljivih sustava za razmjenu informacija o dobavljačima i/ili putem programa na razini sektora.

Napomena: * Primjeri uključuju propise o upravljanju rizicima u EU-u, kao što su Akt o digitalnim uslugama, Direktiva o dužnoj pažnji za održivo poslovanje i Akt o UI-ju.

1. Uključiti uvjete i očekivanja u pogledu tematskih područja OPP-a u ugovore s dobavljačima, prodajnim partnerima i/ili korisnicima ili u druge oblike pisanih sporazuma (npr. izrada „smjernica za odgovornu upotrebu" ili „politika prihvatljive upotrebe" za korisnike).
2. Poticati strane u poslovnim odnosima koje izazivaju štetne učinke ili im doprinose da se savjetuju s pogođenim ili potencijalno pogođenim dionicima ili njihovim predstavnicima te da se uključe u izradu i provedbu korektivnih akcijskih planova.
3. Podupirati mjerodavne strane u poslovnim odnosima u sprečavanju ili ublažavanju štetnih učinaka ili rizika, primjerice osposobljavanjem ili jačanjem njihovih sustava upravljanja, težeći stalnom poboljšanju na osnovi mjerljivih ciljeva s definiranim rokovima.
4. Poticati mjerodavna tijela u zemlji u kojoj je štetni učinak nastao da djeluju, primjerice putem inspekcija te provedbe i primjene postojećih zakona i propisa.
5. Uključiti druga poduzeća i dionike radi zaustavljanja štetnih učinaka i/ili sprečavanja njihova ponovnog nastajanja ili radi sprečavanja materijalizacije rizika (npr. sudjelovanjem u strukovnim inicijativama i uključivanjem vlada).
6. Ako informacije o dubinskoj analizi strane u poslovnim odnosima nisu javno dostupne, nastojati uključiti strane u poslovnim odnosima radi povećanja transparentnosti ili dokazivanja dubinske analize putem povjerljivih bilateralnih ili multilateralnih aranžmana (npr. objavljivanje informacija pouzdanim strukovnim inicijativama ili onima koje obuhvaćaju više dionika, odnosno sklapanje sporazuma o povjerljivosti).
7. Kao krajnju mjeru, razmotriti raskid poslovnih odnosa (v. 2.15.).

OKVIR 2.15. RAZUMIJEVANJE RASKIDA POSLOVNIH ODNOSA U KONTEKSTU RIZIKA

Dinamika između strana u poslovnim odnosima u razvoju i upotrebi sustava umjetne inteligencije stalno se mijenja te su potrebna dodatna istraživanja i savjetovanja kako bi se u potpunosti razumjele implikacije raskida poslovnih odnosa u tom kontekstu. U nekim slučajevima raskid poslovnih odnosa možda neće biti moguć. Umjetna se

inteligencija sve više ugrađuje u osobne i poslovne alate te postaje temeljni dio poslovanja u mnogim sektorima gospodarstva. Isto tako, u nekim fazama razvoja sustava umjetne inteligencije postoji tek mali broj poduzeća, kao što su ona koja razvijaju umjetnu inteligenciju za opće namjene te proizvođači poluvodiča koji predstavljaju ključne izvore opskrbe te je teško ili nemoguće raskinuti poslovne odnose s njima.

Općenito se, u skladu sa Smjernicama za multinacionalna poduzeća, raskid poslovnih odnosa smatra krajnjom mjerom, nakon neuspješnih pokušaja ublažavanja ili kad poduzeće procijeni da ublažavanje nije izvedivo, odnosno radi ozbiljnosti štetnog učinka. Ako postoji mogućnost da poduzeća nastave odnos i pokažu realne izglede za nastavak tog odnosa ili stvarno poboljšanje tijekom vremena, takav pristup često je poželjniji od prekida odnosa.

Stoga su neki od čimbenika koje treba razmotriti pri određivanju odgovarajuće mjere u takvim situacijama utjecaj poduzeća na predmetni subjekt te ozbiljnost učinka. Još neki od mjerodavnih čimbenika jesu i mjera u kojoj je odnos važan za poduzeće te mogući štetni učinci na društvo, okoliš i gospodarstvo povezani s odlukom o raskidu poslovnih odnosa.

U slučajevima u kojima raskid nije moguć preporučuje se da poduzeća interno izvijeste o situaciji, nastave pratiti poslovni odnos, na primjer održavanjem baze podataka znanja, te ponovno razmotre svoju odluku o nastavku poslovnog odnosa ako se okolnosti promijene ili kao dio dugoročne strategije poduzeća da sistemski reagira na sve štetne učinke.

Osim toga, u interesu poduzeća može biti da javno obrazloži odluku o neraskidanju poslovnog odnosa, način na koji se odluka uklapa u politike i prioritete poduzeća, mjere koje poduzima kako bi pokušalo primijeniti svoj utjecaj i na taj način ublažiti štetne učinke, te plan nastavka praćenja poslovnog odnosa u budućnosti.

OKVIR 2.16. PRAKTIČNI PRIMJERI DUBINSKE ANALIZE ZA INVESTITORE I FINANCIJSKE INSTITUCIJE KOJI ULAŽU U RAZVOJ SUSTAVA UMJETNE INTELIGENCIJE

Investitori i financijske institucije imaju ključnu ulogu u razvoju sustava umjetne inteligencije, a poduzeća iz područja umjetne inteligencije privlače znatna financijska sredstva. Tako je globalna godišnja vrijednost ulaganja rizičnog kapitala u umjetnu inteligenciju znatno porasla, s približno 6,4 milijarde USD 2012. na 147 milijardi USD 2024., što čini 56 % vrijednosti svih ulaganja rizičnog kapitala do trećeg tromjesečja 2025. (OECD, n.d.[29]). Neki investitori, osobito oni u ranim fazama ulaganja, često su stoga u poziciji ostvarivanja utjecaja jer podržavaju subjekte u koje se ulaže pri oblikovanju i određivanju smjera projekta.* Još jedno područje u kojem investitori i financijske institucije mogu uvelike utjecati na smjer kretanja jest pružanje smjernica o načinu na koji subjekti u koje se ulaže i klijenti mogu sami ulagati u sustave umjetne inteligencije. Investitori i financijske institucije mogu potaknuti subjekte u koje se ulaže na to da prepoznaju stvarne ili potencijalne štetne učinke i pokušaju pronaći rješenja za njih. U nastavku se navode neki od praktičnih primjera provedbe:

- uključivanje rizika od štetnih učinaka u procjene rizika portfelja ili investicijske analize
- uključivanje dionika u svrhu lakšeg prepoznavanja rizika povezanih s društvima u koja se ulaže
- bilateralni razgovori s društvima u koja se ulaže u svrhu njihova upoznavanja s problematikom utvrđenom u okviru procjene rizika (kao i uključivanja dionika), informiranja o njihovoj praksi dubinske analize te pribavljanja dodatnih informacija i zahtijevanja dodatnih mjera od njih
- zahtijevanje od društava u koja se ulaže jasnih i sažetih obrazloženja za uvođenje sustava umjetne inteligencije
- sudjelovanje u koordiniranim zajedničkim nastojanjima investitora

da uspostave snažniji i usklađeniji pristup odgovornoj umjetnoj inteligenciji
- uspostava inicijativa ili uključivanje u inicijative radi poticanja ili razvoja dobrih praksi / unaprijeđenih tržišnih standarda društava povezanih s odgovornom umjetnom inteligencijom
- potpisivanje javnih obveza s investitorima po pitanjima tematskih područja OPP-a
- prema potrebi, predlaganje odluka dioničara u vezi s pokretanjem pitanja rizika od štetnih učinaka
- ako se ostali načini unapređenja dubinske analize u društvima u koja se ulaže pokažu bezuspješnima, razmotriti glasanje protiv njihovih članova uprave ili, ako je to u skladu s mandatima, prodaju udjela
- objavljivanje javnih priopćenja pri prodaji udjela u društvu ili isključivanju društva iz investicijskog portfelja zbog neprovođenja dubinske analize.

Napomena: * Investitori okupljeni pod okriljem Koalicije za zajedničko djelovanje u području etičke umjetne inteligencije od 2022. provode aktivnosti informiranja i uključivanja usmjerene na 44 od 150 poduzeća obuhvaćenih Indeksom digitalne uključenosti za 2021. Svjetskog saveza za vrednovanje s naglaskom na poduzeća koja nisu objavila skup načela za usmjeravanje razvoja i upotrebe umjetne inteligencije. Te su aktivnosti ponukale dodatnih 14 poduzeća da objave načela za umjetnu inteligenciju u Indeksu digitalne uključenosti za 2023., čime je ukupan broj poduzeća s objavljenim načelima porastao na 47 od 150 (31 %) (World Benchmarking Alliance, 2023[30]).

Izvor: OECD-ov Centar za odgovorno poslovno ponašanje radi na operacionalizaciji dubinske analize odgovornog poslovnog ponašanja za različite financijske transakcije i aktere, izrađujući smjernice prilagođene svrsi za institucionalne ulagače, korporativno kreditiranje i preuzimanje rizika pri izdavanju vrijednosnih papira te financiranje projekata i imovine. OECD (2022[31]), Responsible business conduct due diligence for project and asset finance transactions, https://doi.org/10.1787/952805e9-en.

4. KORAK – PRATITI PROVEDBU I REZULTATE MJERA DUBINSKE ANALIZE

Pratiti provedbu i učinkovitost aktivnosti dubinske analize poduzeća, odnosno mjera koje provodi kako bi prepoznalo, spriječilo i ublažilo štetne učinke te u slučaju potrebe pomoglo njihovu otklanjanju.

TABLICA 2.5. – 4. KORAK: PREGLED POVEZANIH ODREDBI U POSTOJEĆIM OKVIRIMA

ASEAN-ov vodič	Odjeljak C.3 i Prilog A:3
Australske smjernice za uvođenje umjetne inteligencije (Prakse provedbe)	Praksa provedbe 1, 2, 3, 4, 5
Kanadski kodeks ponašanja	Ljudski nadzor i praćenje
HUDERIA	Iterativni zahtjevi
Akt EU-a o UI-ju	Uvodna izjava 114., čl. 9. st. 5. – 8.: Testiranje; čl. 55. st. 1. t. (c): Dostavljanje informacija o incidentima za modele umjetne inteligencije opće namjene; čl. 60.: Testiranje visokorizičnih UI sustava u stvarnim uvjetima izvan regulatornih izoliranih okruženja za umjetnu inteligenciju; čl. 72.: Praćenje nakon stavljanja na tržište
DSA	Čl. 37.: Neovisna revizija
CSDDD	Čl. 14.: Uspostava i održavanje mehanizma za prijave i postupanje s pritužbama; čl. 15.: Praćenje djelotvornosti politike i mjera dužne pažnje
	Načela 4.
ISO 31000 i ISO/IEC 23894	6.6.: Praćenje i preispitivanje
ISO/IEC 42001	8.1.: Operativno planiranje i nadzor; 9.: Vrednovanje uspješnosti; 10.: Poboljšavanje; Prilog A.6.2.6. i A.6.2.8. (Praćenje i evidentiranje); Prilog A.8.4. (Komunikacija o incidentima)
Korejski Osnovni zakon o umjetnoj inteligenciji	Čl. 32. i 34.

Singapurski okvir za testiranje AI Verify	Sigurnost 4.1.1. – 4.6.1.; Zaštita 5.1.1. – 5.7.1.; Robusnost 6.1.1. – 6.5.3.
Vodeća načela UN-a o poslovanju i ljudskim pravima	Operativno načelo 20.
Okvir za osiguravanje pouzdanosti UI-ja, UK DSIT	5.7.: Revizija usklađenosti; 6.1.3.: Preispitivanje unutarnjeg upravljanja i upravljanja rizicima
Okvir SAD-a za upravljanje rizicima UI-a	Mjera 1, 4

Praktični primjeri provedbe (za sva poduzeća u lancu vrijednosti umjetne inteligencije, skupine 1 – 3)

1. Utvrditi štetne učinke ili rizike koje se možda previdjelo u prethodnim postupcima dubinske analize te ih ubuduće uključiti.
2. Procijeniti postoje li prethodno neprepoznati rizici odnosno jesu li prethodno procijenjeni rizici postali neprihvatljivi.
3. Procijeniti učinkovitost nastojanja da se uključi dionike (npr. ispitati je li uključivanje dionika pravodobno, dostupno, primjereno i sigurno).
4. Uključiti povratne informacije o iskustvima dubinske analize poduzeća kako bi se ubuduće poboljšali proces i ishodi.
5. Nadgledati i pratiti učinkovitost sustava umjetne inteligencije ili kriterije pouzdanosti kvalitativno ili kvantitativno u uvjetima sličnima onima prilikom uvođenja sustava, primjerice:
   a) dokumentirajući testne skupove, pokazatelje i pojedinosti o alatima korištenima tijekom testiranja, evaluacije, verifikacije i validacije (TEVV)
   b) prepoznajući i dokumentirajući mjerljiva poboljšanja ili pogoršanja učinkovitosti na temelju savjetovanja s mjerodavnim akterima umjetne inteligencije i drugim dionicima, uključujući zajednice izložene mogućim štetnim učincima, kao i terenskih podataka o rizicima relevantnima za kontekst i obilježjima pouzdanosti
   c) dokumentirajući i dijeleći informacije o incidentima s dionicima, uključujući zajednice izložene mogućim štetnim učincima, vlade, radnike, predstavnike radnika i sindikate, organizacije civilnog društva i akademsku zajednicu. To je potrebno radi unapređenja sigurnosti, zaštite i pouzdanosti naprednih sustava umjetne inteligencije

   d) dokumentirajući rezultate praćenja pouzdanosti sustava umjetne inteligencije u kontekstu ili kontekstima njegova uvođenja, kao i tijekom njegova cjelokupnog životnog ciklusa, te dijeleći te rezultate sa stručnjacima iz domene i mjerodavnim akterima i dionicima umjetne inteligencije kako bi se provjerilo funkcionira li sustav dosljedno u skladu s namjenom.
6. Nadgledati te pratiti provedbu i učinkovitost internih obveza, aktivnosti i ciljeva organizacije u vezi s dubinskom analizom, primjerice provođenjem periodičnih internih revizija ili revizija trećih strana odnosno revizija postignutih rezultata, kao i priopćavanjem rezultata mjerodavnim razinama unutar organizacije.
7. Periodično ocjenjivati poslovne odnose kako bi se provjerilo provode li se mjere ublažavanja rizika ili potvrdilo da su štetni učinci zaista spriječeni ili ublaženi.

5. KORAK – KOMUNICIRATI MJERE PODUZETE RADI RJEŠAVANJA PITANJA UČINAKA

Priopćiti informacije koje su mjerodavne za vanjske subjekte, a koje se odnose na politike, postupke i aktivnosti provedene radi prepoznavanja i pristupanja rješavanju stvarnih ili potencijalnih štetnih učinaka, uključujući rezultate i ishode tih aktivnosti. Komunikacija može imati različite oblike ovisno o ciljnoj skupini (npr. savjetovanja s dionicima i javna komunikacija putem godišnjih izvješća poduzeća, izvješća o održivosti ili društveno odgovornom poslovanju ili drugih odgovarajućih oblika objavljivanja informacija propisanih zakonodavstvom ili dobrovoljnim inicijativama).

TABLICA 2.6. – 5. KORAK: PREGLED POVEZANIH ODREDBI U POSTOJEĆIM OKVIRIMA

ASEAN-ov vodič	Odjeljak C.4. Interakcija i komunikacija s dionicima i Prilog A:5
Australske smjernice za uvođenje umjetne inteligencije (Prakse provedbe)	Praksa provedbe 1, 2, 3, 4

Kanadski kodeks ponašanja	Transparentnost
HUDERIA	Proces uključivanja dionika
Akt EU-a o UI-ju	Čl. 13.: Transparentnost i pružanje informacija subjektima koji uvode sustav; čl. 53. st. 1. t. (a) i (d); čl. 55. st. 1. t. (c)
DSA	Čl. 42.: Obveze izvješćivanja radi transparentnosti
CSDDD	Čl. 16.: Javna komunikacija o dužnoj pažnji
Kodeks ponašanja Hirošimskog procesa	Načelo 4. i 5.
IEEE 7000	11.: Postupak upravljanja transparentnošću
ISO 31000 i ISO/IEC 23894	6.7: Evidentiranje i izvješćivanje
ISO/IEC 42001	7.4: Komunikacija; Prilog A.6.2.7., A.8.2., A.8.4., A.8.5.
Japanske Smjernice za umjetnu inteligenciju u poslovanju	Dio 2C. Zajednička vodeća načela 6, 7; Prilog 3., 4., 5., B. Opisi „Zajedničkih vodećih načela" u dijelu 2. 6, 7
Korejski Osnovni zakon o umjetnoj inteligenciji	Čl. 28. i 31.
Singapurski okvir za testiranje AI Verify	Transparentnost 1.1.1. – 1.5.1.
Vodeća načela UN-a o poslovanju i ljudskim pravima	Operativno načelo 21.
Okvir za osiguravanje pouzdanosti UI-ja, UK DSIT	4.1.3.: Komunicirati
Okvir SAD-a za upravljanje rizicima UI-a	Upravljanje 4

Praktični primjeri provedbe (za sva poduzeća u lancu vrijednosti umjetne inteligencije, skupine 1 – 3)

1. Javno komunicirati sve mjerodavne informacije o postupcima dubinske analize, uvažavajući povjerljivost poslovnih podataka, pravo tržišnog natjecanja te druga pitanja povezana s konkurentnošću ili sigurnošću.[12] Među informacijama koje treba uključiti jesu:
   a) politike OPP-a, u skladu s 1. korakom uključujući informacije o

   obvezama u okviru mjerodavnih dobrovoljnih inicijativa i njihovoj provedbi
   b) postupci za praćenje značajnih incidenata i pogrešaka, odgovor na njih te oporavak nakon njih
   c) značajni štetni učinci ili rizici koji su prepoznati, određeni kao prioritetni i procijenjeni. Kad je riječ o učincima ili drugim značajnim rizicima koje poduzeće uzrokuje ili kojima doprinosi u području ljudskih prava, pravodobno, kulturološki osjetljivo i pristupačno priopćavati pogođenim ili potencijalno pogođenim dionicima sve informacije koje su za njih mjerodavne, pogotovo ako su oni ili netko drugi u njihovo ime izrazili zabrinutost s time u vezi
   d) kriteriji i postupci utvrđivanja prioriteta rizika
   e) mjere koje su poduzete ili ih se planira poduzeti radi sprečavanja ili ublažavanja rizika, ako je moguće, uključujući i očekivane vremenske okvire i referentne vrijednosti za poboljšanja te ishodi, kao što su pojedinosti o provedenim evaluacijama (uključujući penetracijska testiranja) u vezi s rizicima štetnih učinaka
   f) mjere za praćenje provedbe i rezultata
   g) osiguravanje otklanjanja štete ili suradnja na njezinu otklanjanju
   h) sposobnosti, ograničenja te područja primjerene i neprimjerene upotrebe sustava umjetne inteligencije
   i) smislene informacije o svim novim značajnim izdanjima sustava umjetne inteligencije koja mogu ući u široku primjenu
   j) način na koji se mjerodavni dionici uključuju u osmišljavanje i provedbu tih postupaka dubinske analize
   k) ako je mjerodavno (npr. za skupinu 2), informacije o incidentima i pokušajima aktera umjetne inteligencije da zaobiđu zaštitne mehanizme, osobito incidentima povezanima sa sustavima umjetne inteligencije opće namjene.
2. Navedene informacije objavljivati primjereno korisnicima, redovito, pravodobno, pri čemu informacije trebaju biti pouzdane, jasne, potpune, točne i dovoljno detaljne (za dodatne pojedinosti v. Smjernice za multinacionalna poduzeća, poglavlje III.: Objavljivanje informacija).
3. Osigurati da se informacije prikazuju na način primjeren različitim ciljnim skupinama te, prema potrebi, poduzeti posebne mjere kako bi informacije bile dostupne ranjivim dionicima (npr. radnicima).

6. KORAK – OTKLANJATI UČINKE ILI, U SLUČAJU POTREBE, SURAĐIVATI NA NJIHOVU OTKLANJANJU

Ako je poduzeće izazvalo stvarne štetne učinke ili im doprinijelo, nastojati omogućiti povratak pogođene osobe ili više njih na situaciju u kojoj bi bili da nije nastupio štetni učinak (kad je to moguće) te omogućiti otklanjanje proporcionalno značaju i razmjeru štetnog učinka.

TABLICA 2.7. – 6. KORAK: PREGLED POVEZANIH ODREDBI U POSTOJEĆIM OKVIRIMA

ASEAN-ov vodič	Odjeljak C.4.: 4. Interakcija i komunikacija s dionicima
Australske smjernice za uvođenje umjetne inteligencije (Prakse provedbe)	Praksa provedbe 1. i 2.
DSA	Čl. 14.: Uvjeti poslovanja
CSDDD	Čl. 12.: Osiguravanje otklanjanja štete stvarnih negativnih učinaka
Singapurski okvir za testiranje AI Verify	Transparentnost 1.4.1. – 1.5.1.
Vodeća načela UN-a o poslovanju i ljudskim pravima	Operativna načela 22, 29, 30 i 31
Okvir za osiguravanje pouzdanosti UI-ja, UK DSIT	5.3.: Osiguravanje podataka, modela, sustava i upravljanja u praksi

Praktični primjeri provedbe (za sva poduzeća u lancu vrijednosti umjetne inteligencije, skupine 1 – 3)

1. Prema potrebi (tj. u slučajevima kad poduzeće izaziva štetni učinak ili mu doprinosi) uspostaviti legitimne mehanizme za otklanjanje štetnih učinaka ili suradnju s takvim mehanizmima u okviru kojih pogođeni dionici mogu podnositi pritužbe i tražiti

njihovo rješavanje u suradnji s poduzećem (v. okvir 2.17.).
2. Kad je riječ o štetnim učincima izravno povezanima sa stranama u poslovnim odnosima, od poduzeća se očekuje da u okviru prava tržišnog natjecanja primijene svoj utjecaj na te poslovne odnose kako bi ih potaknula na uspostavu mehanizama za otklanjanje štetnih učinaka ili suradnju s takvim mehanizmima.

OKVIR 2.17. MOGUĆE OPCIJE ZA OTKLANJANJE ŠTETNIH UČINAKA

Krajnji cilj otklanjanja jest ponovno uspostavljanje situacije u kojoj bi se oštećenik ili oštećenici nalazili da nije nastupio štetni učinak. Djelotvorno otklanjanje ovisi o kontekstu i postoji niz mogućih opcija. Može se primijeniti više opcija otklanjanja štetnih učinaka, ovisno o razmjeru i kontekstu nastale štete. Te opcije u širem smislu su sljedeće:

- restitucija: povratak pogođene osobe ili skupine osoba u položaj u kojem bi bili da šteta nije nastupila
- naknada štete: financijska naknada za štetu koja se može ekonomski procijeniti, što može uključivati naknadu za tjelesnu ili duševnu štetu, izgubljenu zaradu ili troškove stručne pomoći, poput medicinskih troškova
- rehabilitacija: pružanje medicinske i psihološke skrbi te pravnih i socijalnih usluga
- zadovoljština: utvrđivanje činjenica, odluke usmjerene na vraćanje dostojanstva pogođenim osobama i skupinama, priznavanje nepoštivanja ljudskih prava, isprike, pravne sankcije protiv osobe odgovorne za štetu, kao što su novčane i zatvorske kazne, komemoracije
- jamstva neponavljanja: mjere koje doprinose sprečavanju u budućnosti, uključujući ukidanje sustava, sudske naloge i izmjene korporativnih politika, postupaka i strategija.

Mehanizmi koji mogu olakšati provedbu otklanjanja štetnih učinaka mogu imati različite oblike, među kojima su:

- izravno djelovanje poduzeća (uz savjetovanje s dionicima), bez upućivanja na mehanizam rješavanja sporova
- algoritamske revizije koje provode neovisni, multidisciplinarni paneli

- sudski mehanizmi: nacionalni i regionalni sudovi
- državni izvansudski mehanizmi: mehanizmi povezani s državom koji bi mogli osigurati neki oblik otklanjanja štetnih učinaka, kao što su nacionalne kontaktne točke za OPP, pučki pravobranitelji, inspekcijska tijela i nacionalne institucije za ljudska prava
- izvansudski mehanizmi za podnošenje pritužbi izvan razine države: mehanizmi za otklanjanje štetnih učinaka koje osmišljavaju i kojima upravljaju privatni subjekti, kao što su trgovačka društva ili, u nekim slučajevima, strukovne udruge ili skupine koje uključuju više dionika (v. Vodeće načela UN-a o poslovanju i ljudskim pravima 31 za dodatne informacije o tome što mehanizam otklanjanja štetnih učinaka čini djelotvornim (UN OHCHR, 2021[32])).

Izvor: UN OHCHR B-Tech Project (2021[33]), Access to remedy and the technology sector: basic concepts and principles, https://www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/access-to-remedy-concepts-and-principles.pdf.

LITERATURA

BSR (2022), *Sales partners and human rights due diligence in the technology sector: Best practices brief*, https://www.bsr.org/reports/Sales_Partner_-_Best_Practice_Brief.pdf. [27]

Centre for the Governance of AI (2023), *Towards best practices in AGI safety and governance: A survey of expert opinion*, https://arxiv.org/pdf/2305.07153. [19]

Coalition for Content Provenance and Authenticity (C2PA) (n.d.), *Guiding Principles for C2PA Designs and Specifications*, https://c2pa.org/principles/. [22]

European Center for Not-for-Profit Law (2023), *Framework for Meaningful Engagement*, https://oecd.ai/en/catalogue/tools/framework-for-meaningful-engagement-of-external-stakeholders-in-ai-development. [38]

Europska unija (2024), *Uredba (EU) 2024/1689 Europskog parlamenta i Vijeća od 13.lipnja 2024. o utvrđivanju usklađenih pravila o umjetnoj inteligenciji i o izmjeni uredaba (EZ) br. 300/2008, (EU) br. 167/2013, (EU) br. 168/2013, (EU) 2018/858, (EU) 2018/1139 i (EU) 2019/2144 te direktiva 2014/90/EU, (EU) 2016/797 i (EU) 2020/1828 (Akt o umjetnoj inteligenciji)*, https://eur-lex.europa.eu/legal-content/HR/ALL/?uri=CELEX:32024R1689. [14]

G7 (2023), *Međunarodni kodeks ponašanja u okviru Hirošimskog procesa za napredne sustave UI-ja*, https://www.soumu.go.jp/hiroshimaaiprocess/en/documents.html. [15]

Institute for AI Policy and Strategy (IAPS) (2023), *Deployment corrections: An incident response framework for frontier AI*, https://www.iaps.ai/research/deployment-corrections. [24]

ISO (2023), *ISO/IEC 27035-1:2023: Information security incident management*, https://www.iso.org/standard/78973.html. [25]

ISO (2018), *ISO 31000:2018 Risk management – Guidelines*, https://www.iso.org/standard/65694.html. [36]

MOR (2023.), *Tripartite Declaration of Principles concerning Multinational Enterprises and Social Policy [Tripartitna deklaracija Međunarodne organizacije rada o načelima koja se odnose na multinacionalna poduzeća i socijalnu politiku]*, https://www.ilo.org/publications/tripartite-declaration-principles-concerning-multinational-enterprises-and-3. [5]

NIST (2024), *NIST AI 600-1 Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, https://doi.org/10.6028/NIST.AI.600-1. [41]

OECD (2025), *RBC Spotlight: Due diligence in electronics and vehicle manufacturing*, https://www.oecd.org/en/publications/responsible-business-conduct-spotlights_03a75bf9-en/due-diligence-in-electronics-and-vehicle-manufacturing_993bb959-en.html. [39]

OECD (2025), "Towards a common reporting framework for AI incidents", *OECD Artificial Intelligence Papers*, No. 34, OECD Publishing, Pariz, https://doi.org/10.1787/f326d4ac-en. [28]

OECD (2024), "AI, data governance and privacy: Synergies and areas of international co-operation", *OECD Artificial Intelligence Papers*, No. 22, OECD Publishing, Pariz, https://doi.org/10.1787/2476b1a4-en. [20]

OECD (2024), "Defining AI incidents and related terms", *OECD Artificial Intelligence Papers*, No. 16, OECD Publishing, Pariz, https://doi.org/10.1787/d1a8d965-en. [34]

OECD (2024), *Explanatory memorandum on the updated OECD definition of an AI system*, https://www.oecd.org/en/publications/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_623da898-en.html. [2]

OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", *OECD Artificial Intelligence Papers*, No. 8, OECD Publishing, Pariz, https://doi.org/10.1787/623da898-en. [6]

OECD (2024), *Recommendation of the Council on Artificial Intelligence*, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449. [9]

OECD (2023), "Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI", *OECD Digital Economy Papers*, No. 349, OECD Publishing, Pariz, https://doi.org/10.1787/2448f04b-en. [17]

OECD (2023), *Common Guideposts to Promote Interoperability in AI Risk Management*, https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/11/common-guideposts-to-promote-interoperability-in-ai-risk-management_9629ed36/ba602d18-en.pdf. [10]

OECD (2023), *Smjernice OECD-a za multinacionalna poduzeća o odgovornom poslovnom ponašanju*, OECD Publishing, Pariz, https://doi.org/10.1787/81f92357-en. [1]

OECD (2022), *Measuring environmental impacts of AI compute and applications*, https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/11/measuring-the-environmental-impacts-of-artificial-intelligence-compute-and-applications_3dddded5/7babf571-en.pdf. [7]

OECD (2022), "OECD Framework for the Classification of AI systems", *OECD Digital Economy Papers*, No. 323, OECD Publishing, Pariz, https://doi.org/10.1787/cb6d9eca-en. [13]

OECD (2022), *Report on the Implementation of the Recommendation of the OECD Council on Due Diligence Guidance for Responsible Supply Chains of Minerals from Conflict-Affected and High-Risk Areas*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0386. [37]

OECD (2022), *Responsible Business Conduct Due Diligence for Project and Asset Finance Transactions*, https://doi.org/10.1787/95280529-en. [31]

OECD (2018), *Smjernice OECD-a za dubinsku analizu odgovornog poslovnog ponašanja*, Ministarstvo gospodarstva Republike Hrvatske, 2024., https://rbcroatia.gov.hr/wp-content/uploads/2024/04/Smjernice-OECD-a-za-dubinsku-analizu-1.pdf. [3]

OECD (2016), *OECD Due Diligence Guidance for Responsible Supply Chains of Minerals from Conflict-Affected and High-Risk Areas* [Smjernice za dubinsku analizu za odgovorne lance opskrbe mineralima iz sukobima pogođenih i visokorizičnih područja]: treće izdanje, OECD Publishing, Pariz, https://doi.org/10.1787/9789264252479-en. [40]

OECD (2015), *Competition Law and Responsible Business Conduct*, https://mneguidelines.oecd.org/global-forum/2015GFRBC-Competition-Law-RBC.pdf. [11]

OECD (2015), *Competition Law and Responsible Business Conduct*, https://mneguidelines.oecd.org/global-forum/2015GFRBC-Competition-Law-RBC.pdf. [16]

OECD (2013), *Guidelines governing the protection of privacy and transborder flows of personal data*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188. [18]

OECD (n.d.), *Catalogue of tools and metrics for trustworthy AI*, https://oecd.ai/en/catalogue/tools. [12]

OECD (n.d.), *Worldwide investments in AI*, https://oecd.ai/en/data?selectedArea=investments-in-ai-and-data (pristupljeno 1. listopada 2025.). [29]

Partnership on AI (2021), *Responsible Sourcing of Data Enrichment Services*, https://partnershiponai.org/wp-content/uploads/2021/08/PAI-Responsible-Sourcing-of-Data-Enrichment-Services.pdf. [8]

Partnership on AI (n.d.), *Responsible Practices for Synthetic Media*, https://synthetic-media.partnershiponai.org/#read_the_framework. [23]

Solaiman, I. (2023), *The Gradient of Generative AI Release: Methods and Considerations*, https://arxiv.org/pdf/2302.04844. [26]

UN OHCHR (2021), *OHCHR Accountability and Remedy Project: Meeting the UNGPs' Effectiveness Criteria*, https://www.ohchr.org/sites/default/files/2022-01/arp-note-meeting-effectiveness-criteria.pdf. [32]

UN OHCHR B-Tech Project (2021), *Access to remedy and the technology sector: Basic concepts and principles*, https://www.ohchr.org/sites/default/files/Documents/Issues/Business/B-Tech/access-to-remedy-concepts-and-principles.pdf. [33]

Ured visokog povjerenika Ujedinjenih naroda za ljudska prava (2012), *Vodeća načela Ujedinjenih naroda o poslovanju i ljudskim pravima*, https://www.ohchr.org/en/publications/reference-publications/guiding-principles-business-and-human-rights. [4]

US Department of Commerce. National Telecommunications and Information Administration (2024), *AI Accountability Policy Report*, https://www.ntia.gov/sites/default/files/publications/ntia-ai-report-final.pdf. [21]

US National Institute of Standards and Technology (2023), *AI Risk Management Framework* 1.0, https://doi.org/10.6028/NIST.AI.100-1. [35]

World Benchmarking Alliance (2023), *Digital Inclusion Collective Impact Coalition 2023 Progress Report*, https://archive.worldbenchmarkingalliance.org/impact/digital-inclusion-collective-impact-coalition-progress-report/. [30]

POJMOVNIK

definirati / prepoznati / mapirati / odrediti obuhvat

Pojam u smislu OECD-ovih standarda i izvješća na koje se upućuje u ovim Smjernicama	Okvir interoperabilnosti* preporučuje „definiranje" opsega i konteksta sustava umjetne inteligencije te kriterija za procjenu rizika, primjerice na razini upravljanja, procesa i/ili tehničkoj razini. Ekvivalentni pojam koji se koristi u Smjernicama OECD-a za dubinsku analizu jest „prepoznavanje", a u biti se odnosi na sveobuhvatno određivanje obuhvata radi utvrđivanja svih područja poslovanja, uključujući poslovanje, proizvode i usluge, kao i odnose u kojima je najvjerojatnije da će rizici postojati i biti značajni.
Pojmovi u smislu drugih najvažnijih okvira za upravljanje rizicima	Drugi okviri za upravljanje rizicima, kao što su NIST RMF i ISO 31000, nazivaju taj proces „mapiranje", odnosno „određivanje obuhvata". Slično Okviru interoperabilnosti, NIST RMF iznosi konkretne preporuke o tome što je potrebno mapirati, primjerice namjenu i načine upotrebe sustava umjetne inteligencije, rizike povezane s njegovim sastavnicama, uključujući softver i podatke koje koristi, kao i njegove učinke na pojedince, skupine i društvo. Vodeća načela UN-a o poslovanju i ljudskim pravima usklađena su sa Smjernicama za multinacionalna poduzeća i Smjernicama OECD-a za dubinsku analizu (v. načelo 15.).
Razlike u terminologiji i primjeni u ovim Smjernicama	Različiti instrumenti rabe pojmove „definirati", „prepoznati", „mapirati" i „odrediti obuhvat", općenito podrazumijevajući isti sveobuhvatni skup mjera. Čini se da je funkcija „MAPIRANJE" u okviru NIST RMF-a najkonkretnija i najbolje primjenjiva na sustave umjetne inteligencije. Isto tako, svi okviri za upravljanje rizicima opisuju „definiranje" (odnosno „prepoznavanje" / „mapiranje" / „određivanje obuhvata") kao zaseban korak u procesu upravljanja rizicima koji je ključan kao temelj za procjenu rizika i učinaka. Za potrebe ovih Smjernica smatra se da pojam „DEFINIRATI" obuhvaća sve te aspekte.

dionici

Pojam u smislu OECD-ovih standarda i izvješća na koje se upućuje u ovim Smjernicama	U OECD-ovoj Preporuci o umjetnoj inteligenciji dionici su definirani kao sve organizacije i pojedinci koji su izravno ili neizravno uključeni u sustave umjetne inteligencije ili izloženi njihovu utjecaju. OECD-ove Smjernice za multinacionalna poduzeća navode da su relevantni dionici osobe ili skupine, ili njihovi ovlašteni predstavnici, koji imaju prava ili interese u vezi s područjima obuhvaćenima Smjernicama za multinacionalna poduzeća na koje utječu ili bi mogli utjecati štetni učinci povezani s poslovanjem, proizvodima ili uslugama.

Pojam u smislu drugih okvira za upravljanje rizicima	Drugi okviri rabe pojam dionici kao opći naziv za organizacije i pojedince. U različitim kontekstima „mjerodavni dionici" se obično odnose na korisnike sustava umjetne inteligencije, civilno društvo, predstavnike radnika, mala i srednja poduzeća te druga poduzeća.
Razlike u terminologiji i primjeni u ovim Smjernicama	Za potrebe ovih Smjernica dionike u najširem smislu treba razumjeti kao osobe, skupine ili organizacije koje su uključene u sustave umjetne inteligencije ili na koje oni utječu, kao i poduzeća uključena u njihov razvoj i upotrebu.

dubinska analiza / dužna pažnja

Pojam u smislu OECD-ovih standarda i izvješća na koje se upućuje u ovim Smjernicama	Dubinska analiza odgovornog poslovnog ponašanja u Smjernicama za multinacionalna poduzeća definira se kao postupak koji, kao sastavni dio sustava donošenja poslovnih odluka i upravljanja rizicima, poduzećima omogućava da identificiraju, spriječe i ublaže stvarne i moguće štetne učinke svojih aktivnosti te pojasne kako pristupaju rješavanju tog pitanja. Smjernice za multinacionalna poduzeća obuhvaćaju sljedeće mjere u okviru postupka dubinske analize: 1. uključivanje OPP-a u politike i sustave upravljanja, 2. identificiranje i procjena stvarnih i potencijalnih štetnih učinaka povezanih s poslovanjem, proizvodima ili uslugama poduzeća, 3. zaustavljanje, sprečavanje i ublažavanje štetnih učinaka, 4. praćenje provedbe i rezultata, 5. obavještavanje o načinu svladavanja učinaka te 6. pružanje pomoći ili suradnja u otklanjanju štetnih učinaka, prema potrebi.
Pojam u smislu drugih okvira za upravljanje rizicima	Vodeća načela Ujedinjenih naroda o poslovanju i ljudskim pravima (Vodeća načela UN-a), izrađena usporedno s ažuriranjem Smjernica za multinacionalna poduzeća iz 2011., usklađena su s njima u pogledu pojma dubinske analize. Vodeća načela UN-a definiraju dubinsku analizu kao proces koji uključuje „procjenu stvarnih i potencijalnih učinaka na ljudska prava, integraciju i djelovanje na temelju rezultata procjene, praćenje odgovora i priopćavanje načina na koji se pristupa pitanju učinaka" (OECD, 2023[17]). Vodeća načela UN-a također proširuju očekivanja u pogledu dubinske analize na učinke koje poduzeća uzrokuju, kojima doprinose ili s kojima su izravno povezana putem svojih poslovnih odnosa. Drugi okviri za upravljanje rizicima obično upućuju na definiciju „upravljanja rizicima" iz norme ISO 31000 ili je prilagođavaju kako bi obuhvatili procese slične očekivanjima u pogledu dubinske analize iz Smjernica za multinacionalna poduzeća ili povezane s tim očekivanjima. Prema normi ISO 31000 „[u]pravljanje rizicima predstavlja koordinirane aktivnosti za usmjeravanje i nadzor organizacije s obzirom na rizik."

Razlike u terminologiji i primjeni u ovim Smjernicama	Općenito su pojmovi povezani s dubinskom analizom u različitim okvirima usklađeni. Ključna terminološka razlika jest u tome da se u Smjernicama za multinacionalna poduzeća i Vodećim načelima UN-a izričito upućuje na dubinsku analizu koja se odnosi na strane u poslovnim odnosima poduzeća, umjesto usmjerenosti isključivo na poslovanje, proizvode i usluge samog poduzeća. Za potrebe ovih Smjernica dubinska analiza može se razumjeti kao širi pojam koji se koristi u Smjernicama za multinacionalna poduzeća / Smjernicama OECD-a za dubinsku analizu i Vodećim načelima UN-a, koji obuhvaća upravljanje rizicima u poslovanju, proizvodima i uslugama same organizacije, kao i strana s kojima je u poslovnim odnosima.

postupati / zaustaviti, sprečavati, ublažavati i otklanjati / upravljati

Pojam u smislu OECD-ovih standarda i izvješća na koje se upućuje u ovim Smjernicama	Okvir interoperabilnosti definira „postupanje s rizikom" kao primjenu tehnika za sprečavanje, ublažavanje ili zaustavljanje rizika na temelju njihove vjerojatnosti i učinka. Sličan pristup postupanju s rizikom zagovara OECD u svojim preporukama za dubinsku analizu, dodatno precizirajući vrste mjera koje treba poduzeti, ovisno o odgovornosti poduzeća za nastanak rizika (npr. je li ga prouzročilo, doprinijelo njegovu nastanku ili je s njime izravno povezano). Od poduzeća se očekuje da zaustavljaju, sprečavaju i/ili ublažavaju utvrđene rizike i učinke. U situacijama u kojima poduzeće doprinosi učinku ili je s njime izravno povezano putem poslovnog odnosa, ono bi trebalo, u mjeri u kojoj je to moguće, nastojati iskoristiti svoj utjecaj, samostalno ili u suradnji s drugima, kako bi se postigla promjena. Ako je poduzeće prouzročilo štetni učinak ili mu doprinijelo, od njega se očekuje da osigura otklanjanje tog učinka ili da surađuje na njegovu otklanjanju (v. daljnja razmatranja pojma „OTKLANJANJE" u nastavku).
Pojam u smislu drugih okvira za upravljanje rizicima	NIST RMF u okviru svoje funkcije UPRAVLJANJA postupanje s rizikom opisuje kao „planove odgovora na incidente ili događaje, oporavak od njih te priopćavanje informacija o njima". Iako koristi drukčiju terminologiju, skup mjera u okviru te funkcije u pravilu je usklađen s preporukama OECD-a za dubinsku analizu. Očekivane mjere uključuju određivanje prioriteta i raspodjelu resursa za upravljanje rizicima, uključivanje pogođenih dionika te kontinuirano praćenje i dokumentiranje aktivnosti upravljanja rizicima. NIST RMF isto tako ističe kako mogućnosti odgovora na rizik mogu uključivati ublažavanje, prijenos, izbjegavanje ili prihvaćanje rizika.

Razlike u terminologiji i primjeni u ovim Smjernicama	Za potrebe ovih Smjernica pojam „postupati" može se definirati kao zaustavljanje, sprečavanje ili ublažavanje rizika koje poduzeća uzrokuju, kojima doprinose ili s kojima su izravno povezana putem svojih poslovnih odnosa.

praćenje i preispitivanje / slijeđenje

Pojam u smislu OECD-ovih standarda i izvješća na koje se upućuje u ovim Smjernicama	Pod „praćenjem i preispitivanjem" misli se na kontinuiranu aktivnost provjere rizika i mjera poduzetih radi njihova rješavanja. Ekvivalentni pojam u Smjernicama OECD-a za dubinsku analizu jest „pratiti" (odnosno „slijediti"). Konkretno, riječ je o „praćenju" provedbe, učinkovitosti i ishoda aktivnosti dubinske analize (tj. mjera za prepoznavanje, sprečavanje, ublažavanje i otklanjanje štetnih učinaka), uključujući i u odnosu na strane u poslovnim odnosima. Praćenje se provodi periodično i može uključivati neovisne revizije trećih strana.
Pojmovi u smislu drugih najvažnijih okvira za upravljanje rizicima	Drugi okviri na sličan način upućuju na praćenje i preispitivanje. U nekim slučajevima riječ je o zasebnom koraku u postupku dubinske analize (npr. u normi ISO 31000), a u drugim slučajevima dio je sveobuhvatnije mjere usmjerene na unutarnje upravljanje (npr. u okviru NIST-ova Okvira za upravljanje rizikom (NIST RMF)). Načela UN-a o poslovanju i ljudskim pravima usklađena su sa Smjernicama za multinacionalna poduzeća i Smjernicama OECD-a za dubinsku analizu (v. načela 17. i 20.).
Razlike u terminologiji i primjeni u ovim Smjernicama	Općenito su pojmovi u različitim okvirima istovjetni.

procijeniti / mjeriti / vrednovati/evaluirati

Pojam u smislu OECD-ovih standarda i izvješća na koje se upućuje u ovim Smjernicama	Okvir interoperabilnosti taj korak u procesu opisuje kao prepoznavanje rizika, analizu mehanizama putem kojih ti rizici mogu nastati te procjenu njihove vjerojatnosti i ozbiljnosti. Smjernice OECD-a za dubinsku analizu koriste isti pojam na sličan način, istodobno preporučujući poduzećima da procijene svoju povezanost s učincima koje strane u poslovnim odnosima uzrokuju ili kojima doprinose. U taj je korak uključen i postupak određivanja prioriteta rizika, pri kojem poduzeća određuju kojim će najznačajnijim rizicima i učincima, na temelju njihove ozbiljnosti i vjerojatnosti, dati prednost u pogledu poduzimanja mjera.

Pojam u smislu drugih okvira za upravljanje rizicima	Drugi okviri za upravljanje rizicima, kao što su Okvir SAD-a za upravljanje rizicima UI-ja (NIST RMF), ISO 31000, IEEE 7000-21 i ISO/IEC Guide 51, taj korak u procesu označavaju kao mjerenje ili evaluaciju. U okviru NIST RMF-a mjerenje rizika uključuje pokazatelje obilježja pouzdanosti i društvenog učinka radi praćenja rizika. Vodeća načela UN-a o poslovanju i ljudskim pravima usklađena su sa Smjernicama za multinacionalna poduzeća i Smjernicama OECD-a za dubinsku analizu (v. načelo 17. i 24.).
Razlike u terminologiji i primjeni u ovim Smjernicama	Ti instrumenti koriste navedene pojmove na sličan način, ali im je obuhvat malo drukčiji. Primjerice, procjena može obuhvatiti ciljeve pouzdanosti, uključenost u učinak (npr. uzrokovanje, doprinos ili izravna povezanost) ili rizike za poduzeće. Za potrebe ovih Smjernica pojam „PROCJENJIVANJE" može obuhvaćati sve te aspekte.

* Okvir interoperabilnosti proizlazi iz dokumenta Unapređivanje odgovornosti u području umjetne inteligencije *(Advancing Accountability in AI)* u kojem su utvrđeni glavni zajednički koraci za upravljanje rizicima umjetne inteligencije koji se javljaju u više okvira (OECD, 2023[17]).

rizici / incidenti / opasnosti

Pojam u smislu OECD-ovih standarda i izvješća na koje se upućuje u ovim Smjernicama	U OECD-ovim instrumentima za OPP opisuju se štetni učinci ili potencijalni štetni učinci (tj. rizici) u kontekstu tema obuhvaćenih u poglavljima Smjernica za multinacionalna poduzeća: ljudska prava, uključujući radnike i odnose između poslodavca i radnika, okoliš, podmićivanje i korupcija, objavljivanje informacija te interesi potrošača. Smjernice za multinacionalna poduzeća upućuju na rizike kao na vjerojatnost štetnih učinaka na ljude, okoliš i društvo koje poduzeća prouzroče, kojima doprinose ili s kojima su izravno povezana. Drugim riječima, takav je pristup riziku usmjeren prema van. Vjerojatnost nastanka štetnih učinaka povećava se u situacijama u kojima ponašanje poduzeća ili okolnosti povezane s njihovim lancima opskrbe ili stranama u poslovnim odnosima nisu u skladu s preporukama iz Smjernica za multinacionalna poduzeća. Rizik od štetnih učinaka može postojati kada postoji mogućnost ponašanja koja nije u skladu s preporukama iz Smjernica za multinacionalna poduzeća jer uključuje učinke koji se mogu pojaviti u budućnosti. U zasebnom, ali povezanom području rada na praćenju incidenata povezanih s umjetnom inteligencijom, OECD-ova mreža stručnjaka radi na razvoju definicije incidenata i opasnosti povezanih s umjetnom inteligencijom (OECD, 2024[34]), koji se definiraju na sljedeći način:

	incident povezan s umjetnom inteligencijom jest događaj, okolnost ili niz događaja u kojima razvoj, upotreba ili neispravno funkcioniranje jednog ili više sustava umjetne inteligencije izravno ili neizravno dovodi do neke od sljedećih vrsta štete: (a) ozljeda ili šteta po zdravlje osobe ili skupine osoba, (b) poremećaj upravljanja i funkcioniranja kritične infrastrukture, (c) nepoštovanje ljudskih prava ili kršenje obveza iz primjenjivog prava kojima se štite radna i prava intelektualnog vlasništva, (d) šteta za imovinu, zajednice ili okoliš. Opasnost povezana s umjetnom inteligencijom jest događaj, okolnost ili niz događaja u kojima razvoj, upotreba ili neispravno funkcioniranje jednog ili više sustava umjetne inteligencije može razumno dovesti do incidenta povezanog s umjetnom inteligencijom, odnosno jedne od sljedećih vrsta štete: (a) ozljeda ili šteta po zdravlje osobe ili skupine osoba, (b) poremećaj upravljanja i funkcioniranja kritične infrastrukture, (c) nepoštovanje ljudskih prava ili kršenje obveza iz primjenjivog prava kojima se štite radna i prava intelektualnog vlasništva, (d) šteta za imovinu, zajednice ili okoliš.
Pojmovi u smislu drugih okvira za upravljanje rizicima	Akt EU-a o UI-ju definira rizik kao kombinaciju vjerojatnosti nastanka štete i težine te štete (Europska unija, 2024[14]), pri čemu se „šteta" tumači kao šteta javnim interesima i temeljnim pravima zaštićenima pravom Europske unije. Takva šteta može biti materijalna ili nematerijalna, uključujući tjelesnu, psihološku, društvenu ili ekonomsku štetu. US NIST RMF primjenjuje isti pristup kao norma ISO 31000, promatrajući rizik i u pozitivnom i u negativnom smislu (US National Institute of Standards and Technology, 2023[35]) (ISO, 2018[35]). Rizik definira kao složenu mjeru u okviru koje se sagledava vjerojatnost nastanka događaja i razmjere ili stupanj njegovih posljedica. Učinci ili posljedice sustava umjetne inteligencije mogu biti pozitivni, negativni ili i jedno i drugo te mogu predstavljati priliku ili prijetnju.
Razlike u terminologiji i primjeni u ovim Smjernicama	Definicija rizika u Smjernicama za multinacionalna poduzeća šira je po opsegu jer uključuje učinke „koje poduzeća prouzroče, kojima doprinose ili s kojima su izravno povezana". Načelno se time relevantni rizici i učinci proširuju izvan poslovanja samog poduzeća na učinke i rizike povezane sa stranama u poslovnim odnosima u lancu vrijednosti. Unutar različitih okvira razlikuje se i obuhvat konkretnih rizika.

	Za potrebe ovih Smjernica rizik se može razumjeti u skladu s definicijom iz Smjernica za multinacionalna poduzeća jer obuhvaća incidente i opasnosti povezane s umjetnom inteligencijom, ali je širi za potrebe dubinske analize u lancu vrijednosti. Kad je to mjerodavno i konkretnije, u Smjernicama se upućuje na incidente i opasnosti.

sustav umjetne inteligencije / sustav UI-ja / UI sustav

Pojam u smislu OECD-ovih standarda i izvješća na koje se upućuje u ovim Smjernicama	Definicija sustava umjetne inteligencije preuzeta je iz OECD-ove Preporuke o umjetnoj inteligenciji ažurirane 2023. (OECD, 2024[6]). Preporuka definira sustav umjetne inteligencije kao strojni sustav koji, za eksplicitne ili implicitne ciljeve, iz ulaznih vrijednosti koje prima, zaključuje kako generirati izlazne vrijednosti, kao što su predviđanja, sadržaj, preporuke ili odluke koji mogu utjecati na fizička ili virtualna okruženja. Različiti sustavi umjetne inteligencije razlikuju se prema razini autonomije i prilagodljivosti nakon uvođenja. Razlozi za takvu definiciju detaljno su objašnjeni u memorandumu koji je objavio OECD (2024[2]).
Pojam u smislu drugih okvira za upravljanje rizicima	Akt EU-a o UI-ju, Okvir SAD-a za upravljanje rizicima UI-a, ASEAN-ov vodič i Konvencija Vijeća Europe izravno preuzimaju definiciju iz OECD-ove Preporuke o umjetnoj inteligenciji, upućuju na nju ili tek neznatno odstupaju od nje.
Razlike u terminologiji i primjeni u ovim Smjernicama	Za potrebe ovih Smjernica sustav umjetne inteligencije može se razumjeti u skladu s definicijom iz OECD-ove Preporuke o umjetnoj inteligenciji.

BILJEŠKE

1 Preporuka o umjetnoj inteligenciji utvrđuje pet komplementarnih načela utemeljenih na vrijednostima koja su mjerodavna za sve dionike te pet preporuka za donositelje politika i normi, koji se u ovim Smjernicama zajedno nazivaju „načela umjetne inteligencije".

2 Prema Smjernicama za multinacionalna poduzeća, pogl. I., t. 4., „Za potrebe Smjernica nije potrebna precizna definicija multinacionalnih poduzeća. Iako se Smjernicama omogućuje širok pristup pri utvrđivanju subjekata koji se mogu smatrati multinacionalnim poduzećima za potrebe Smjernica, glavni čimbenici koje je potrebno uzeti u obzir u tom pogledu su međunarodna priroda strukture ili djelatnosti poduzeća te njegov komercijalni oblik, svrha ili aktivnosti."

3 Životni vijek sustava umjetne inteligencije obično obuhvaća nekoliko faza, uključujući: planiranje i osmišljavanje; prikupljanje i obradu podataka; izgradnju jednog ili više modela i/ili prilagodbu postojećih modela specifičnim zadaćama; testiranje, evaluaciju, verifikaciju i validaciju; stavljanje na raspolaganje za upotrebu/uvođenje; rad i praćenje; te povlačenje iz upotrebe / stavljanje izvan pogona. Te se faze često odvijaju iterativno i nisu nužno uzastopne. Odluka o povlačenju sustava umjetne inteligencije iz upotrebe može se donijeti u bilo kojem trenutku tijekom faze rada i praćenja. U Načelima umjetne inteligencije, poduzeća uključena u životni ciklus sustava umjetne inteligencije nazivaju se „akteri umjetne inteligencije" (OECD, 2024[2]).

4 Klasifikacija na temelju aktivnosti osmišljena je u OECD-ovu izvješću o unapređivanju odgovornosti u području umjetne inteligencije *(Advancing accountability in AI*, OECD, 2023[17]) te dodatno razrađena u nacrtu izvješća Nacrt mapiranja i objedinjavanja mjerodavnih aktera, tematskih područja i terminologije za odgovorno poslovno ponašanje u području umjetne inteligencije *(Draft mapping and consolidation of relevant actors, issues, and terminology for Responsible Business Conduct in AI)* [DSTI/CDEP/AIGO(2023)[12], o kojem se raspravljalo na sastancima Radne skupine za upravljanje umjetnom inteligencijom (AIGO) i Radne skupine za odgovorno poslovno ponašanje (WPRBC) u studenome 2022.

5 Za više informacija o pristupanju rizicima povezanima sa sirovinama v. Smjernice OECD-a za dubinsku analizu odgovornih lanaca opskrbe mineralima (OECD, 2016[39]); za više informacija o pristupanju rizicima

povezanima s proizvodnjom elektronike i vozila v. Pregled OPP-a o dubinskoj analizi u proizvodnji elektronike i vozila *(RBC Spotlight on Due Diligence in Electronics and Vehicle Manufacturing)* (OECD, 2025[38]).

6 Pojam „uvođenje" se u Načelima umjetne inteligencije rabi drukčije nego u Aktu EU-a o UI-ju. U okviru Načela umjetne inteligencije uvođenje se može razumjeti kao stavljanje sustava umjetne inteligencije na raspolaganje za upotrebu. Prema Aktu EU-a o UI-ju „subjekt koji uvodi sustav" znači fizička ili pravna osoba, tijelo javne vlasti, javna agencija ili drugo javno tijelo koje upotrebljava UI sustav u okviru svoje nadležnosti, osim ako se UI sustav upotrebljava u osobnoj neprofesionalnoj djelatnosti (Europska unija, 2024[14]). Definicija subjekta koji uvodi sustav u Aktu EU-a o UI-ju bliža je onome što se u ovim Smjernicama opisuje kao skupina 3: korisnici sustava umjetne inteligencije.

7 Nacionalne kontaktne točke (NKT-ovi) ovlaštene su unapređivati učinkovitost Smjernica za multinacionalna poduzeća kroz sljedeće mjere: poduzimanje promotivnih aktivnosti, obradu upita i doprinos rješavanju pitanja koja se javljaju u vezi s provedbom Smjernica za multinacionalna poduzeća u konkretnim slučajevima. Bilo koji pojedinac ili organizacija mogu nacionalnoj kontaktnoj točki iznijeti određeni slučaj (predmet) protiv nekog poduzeća tamo gdje ono posluje ili ima sjedište u pogledu njegova poslovanja bilo gdje u svijetu. Nacionalne kontaktne točke olakšavaju pristup sporazumnim i nekontradiktornim postupcima, kao što su mirenje ili posredovanje, kako bi pomogle strankama u rješavanju problematičnih pitanja. Nacionalne kontaktne točke dužne su po zaključenju konkretnih postupaka objavljivati završne izjave. Nacionalne kontaktne točke isto tako mogu davati preporuke na temelju okolnosti konkretnog slučaja. Za informacije o samom postupku pred nacionalnim kontaktnim točkama, kao i o pojedinim nacionalnim kontaktnim točkama ili predmetima, organizacije se upućuje na Smjernice OECD-a za multinacionalna poduzeća o odgovornom poslovnom ponašanju (OECD, 2025[38]).

8 Smjernice za multinacionalna poduzeća također navode da bi poduzeća „posebnu pozornost trebala posvetiti negativnim učincima na pojedince, na primjer na borce za ljudska prava, koji mogu biti izloženi povećanom riziku zbog marginalizacije, ranjivosti ili drugih okolnosti, pojedinačno ili kao članovi određene skupine ili populacije, uključujući

autohtone narode. Smjernice OECD-a za dubinsku analizu, uključujući Smjernice OECD-a za dubinsku analizu odgovornog poslovnog ponašanja, Smjernice OECD-a o dubinskoj analizi za svrhovitu uključenost dionika u ekstraktivnom sektoru i Smjernice OECD-a/FAO-a za odgovorne lance opskrbe u poljoprivredi pružaju daljnje praktične savjete po tom pitanju, uključujući u pogledu dobrovoljnog prethodnog informiranog pristanka. U instrumentima Ujedinjenih naroda su razrađena prava autohtonih naroda (Deklaracija Ujedinjenih naroda o pravima autohtonih naroda)." (Komentar 45.)

9 Kako bi uključivanje dionika bilo svrhovito, treba biti dvosmjerno, provoditi se u dobroj vjeri i uzimati u obzir stavove dionika. Dionicima treba pružiti istinite i potpune informacije te im pravodobno omogućiti da daju svoj doprinos prije donošenja važnih odluka koje bi mogle imati učinak na njih (v. (OECD, 2018[3])).

10 Za više informacija o tome kako svrhovito uključivati dionike pri osmišljavanju proizvoda i usluga utemeljenih na umjetnoj inteligenciji v. Okvir za svrhovito uključivanje vanjskih dionika u razvoj umjetne inteligencije Europskog centra za pravo neprofitnih organizacija (ECNL) (European Center for Not-for-Profit Law, 2023[37]).

11 Primjeri uključuju OECD-ov alat za praćenje incidenata povezanih s umjetnom inteligencijom (AI Incidents Monitor) (OECD, 2025[28]) i bazu podataka OECD-ovih nacionalnih kontaktnih točaka o konkretnim predmetima (OECD, 2022[36]).

12 Objavljivanje informacija ne bi smjelo stvarati nerazmjerna administrativna ili financijska opterećenja za poduzeća. Osim toga, od poduzeća se ne bi trebalo očekivati da objavljuju informacije koje mogu ugroziti njihov konkurentski položaj, osim ako je objavljivanje takvih informacija potrebno kako bi ulagači bili potpuno informirani pri donošenju odluka te kako bi se izbjeglo njihovo dovođenje u zabludu.

SMJERNICE OECD-A ZA DUBINSKU ANALIZU ODGOVORNE UMJETNE INTELIGENCIJE

HRVATSKA NACIONALNA
KONTAKTNA TOČKA

Ovo izvješće pruža praktične smjernice poduzećima za provedbu OECD-ovih standarda odgovornog poslovnog ponašanja (OPP) i Načela OECD-a o umjetnoj inteligenciji pri razvoju i upotrebi umjetne inteligencije (UI). Cilj mu je poduprijeti inovacije, ulaganja i rast poduzeća u lancu vrijednosti umjetne inteligencije pomažući poduzećima da proaktivno pristupaju pitanju štetnih učinaka. Izvješće promiče usklađenost politika te, gdje je to moguće, OECD-ova okvira i drugih nacionalnih ili međunarodnih okvira za upravljanje rizicima umjetne inteligencije.